

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 18:01:21 ; Search time 6.51797 Seconds  
(without alignments)  
6329.459 Million cell updates/sec

Title: US-09-997-610-1\_COPY\_56\_136

Perfect score: 81  
Sequence: 1 gacctccagcacaccacagc.....cacaatatacagaggaata 81

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 424239 seqs, 25461826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCF\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCFUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81	100.0	1381	9 US-09-997-610-1	Sequence 1, Appl
2	81	100.0	1731	9 US-09-997-610-5	Sequence 5, Appl
3	57	70.4	1377	9 US-09-997-610-3	Sequence 3, Appl
4	57	70.4	1731	9 US-09-997-610-7	Sequence 7, Appl
5	32.4	40.0	392	10 US-09-960-352-1786	Sequence 1786, Ap
6	32.4	40.0	415	10 US-09-960-352-9137	Sequence 9137, Ap
7	30.4	37.5	549	10 US-09-923-779-49	Sequence 49, Appl
8	30.4	37.5	3394	10 US-09-880-107-2178	Sequence 2178, Ap
9	30	37.0	2686	9 US-09-764-868-48	Sequence 48, Appl
10	30	37.0	3880	10 US-09-799-799-1	Sequence 1, Appl
11	29.8	36.8	561	10 US-09-923-779-47	Sequence 47, Appl
12	29.8	36.8	617	10 US-09-923-779-51	Sequence 51, Appl
13	29.8	36.8	648	10 US-09-923-779-66	Sequence 66, Appl
14	29.8	36.8	653	10 US-09-923-779-53	Sequence 53, Appl
15	29.8	36.8	655	10 US-09-923-779-68	Sequence 68, Appl
16	29.8	36.8	656	10 US-09-923-779-69	Sequence 69, Appl
17	29.8	36.8	658	10 US-09-923-779-71	Sequence 71, Appl
18	29.8	36.8	659	10 US-09-923-779-52	Sequence 52, Appl
19	29.8	36.8	780	10 US-09-923-779-5	Sequence 5, Appl

20	29.8	36.8	823	10 US-09-923-779-28	Sequence 28, Appl
21	29.8	36.8	88191	10 US-09-799-799-3	Sequence 3, Appl
22	29.6	36.5	1619	9 US-09-764-868-400	Sequence 400, App
23	29.4	36.3	821	9 US-09-729-6588-7	Sequence 7, Appl
24	29.4	36.3	1176	9 US-09-729-6588-14	Sequence 14, Appl
25	29.4	36.3	1574	9 US-09-729-6588-1	Sequence 1, Appl
26	28.6	35.3	886	10 US-09-923-779-4	Sequence 4, Appl
27	28.6	35.3	6158	10 US-09-919-497-6	Sequence 762, App
28	28.6	35.3	6158	10 US-09-954-456-762	Sequence 762, App
29	28.4	35.1	1320	10 US-09-815-242-7615	Sequence 7615, Ap
30	27.6	34.1	572	10 US-09-923-779-56	Sequence 56, Appl
31	27.6	34.1	684973	10 US-09-263-959-1	Sequence 1, Appl
32	27.2	33.6	274	10 US-09-923-876-4300	Sequence 4300, Ap
33	27	33.3	3913	9 US-10-026-188-7	Sequence 7, Appl
34	27	33.3	155074	9 US-10-026-188-6	Sequence 6, Appl
35	26.8	33.1	259	10 US-09-765-231A-49	Sequence 49, Appl
36	26.8	33.1	2235	10 US-09-962-436-308	Sequence 308, App
37	26.8	33.1	3226	10 US-09-954-456-725	Sequence 725, App
38	26.8	33.1	3690	12 US-10-044-090-448	Sequence 448, App
39	26.2	32.3	1133	9 US-09-924-340-57	Sequence 57, Appl
40	26.2	32.3	1133	9 US-09-992-600A-57	Sequence 57, Appl
41	26.2	32.3	2504	9 US-10-033-297-3	Sequence 3, Appl
42	26.2	32.3	2504	9 US-10-081-806-3	Sequence 3, Appl
43	26.2	32.3	2504	9 US-10-074-328-3	Sequence 3, Appl
44	26.2	32.3	2511	10 US-09-777-430A-9	Sequence 9, Appl
45	26.2	32.3	2511	10 US-09-777-430A-12	Sequence 12, Appl

## ALIGNMENTS

RESULT 1  
US-09-997-610-1  
; Sequence 1, Application US/09997610  
; Patent No. US20020156244A1  
; GENERAL INFORMATION:  
; APPLICANT: Fox, Brian  
; APPLICANT: Holloway, James L.  
; TITLE OF INVENTION: ADIPOCYTE COMPLEMENT RELATED PROTEIN  
; FILE REFERENCE: 00-96  
; CURRENT APPLICATION NUMBER: US/09/997,610  
; CURRENT FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: US 60/253,924  
; PRIOR FILING DATE: 2000-11-29  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 1381  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (2)...(1381)  
US-09-997-610-1

Query Match 100.0%; Score 81; DB 9; Length 1381;  
Best Local Similarity 100.0%; Pred. No. 7.1e-18;  
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 56 GGACCTCCAGCACACCCAGCCCGCCAGAGAGAGTGGGCTCTGTCAGCAGGTTA 60  
OY 1 GGACCTCCAGCACACCCAGCCCGCCAGAGAGAGTGGGCTCTGTCAGCAGGTTA 60  
DB 61 CCACATATATACAGAGATA 81  
OY 116 CCACATATATACAGAGATA 136  
DB 116 CCACATATATACAGAGATA 136

RESULT 2  
US-09-997-610-5  
; Sequence 5, Application US/09997610  
; Patent No. US20020156244A1

```
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 1731
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1731)
US-09-997-610-5
```

```
Query Match          100.0%; Score 81; DB 9; Length 1731;
Best Local Similarity 100.0%; Pred. No. 7.6e-18;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 GGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCGTGACACAGGTTTA 60
    |||
DB 409 GGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCGTGACACAGGTTTA 468
OY 61 CCACATATACAGAGAATA 81
    |||
DB 469 CCACATATACAGAGAATA 489
```

```
RESULT 3
US-09-997-610-3
; Sequence 3, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1377
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide encoding a polypeptide
; OTHER INFORMATION: of SEQ ID NO:2
; NAME/KEY: variation
; LOCATION: (1)...(1377)
; OTHER INFORMATION: Each n is independently A, T, G, or C.
; NAME/KEY: misc_feature
; LOCATION: (1)...(1377)
; OTHER INFORMATION: n = A,T,C or G
US-09-997-610-3
```

```
Query Match          70.4%; Score 57; DB 9; Length 1377;
Best Local Similarity 64.2%; Pred. No. 6.7e-10;
Matches 52; Conservative 9; Mismatches 20; Indels 0; Gaps 0;
```

```
OY 1 GGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCGTGACACAGGTTTA 60
    |||
DB 55 GGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCGTGACACAGGTTTA 114
```

```
OY 61 CCACATATACAGAGAATA 81
    |||
DB 115 CCNCATATACAGAGAATA 135
```

```
RESULT 4
US-09-997-610-7
; Sequence 7, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1731
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide sequence of zacrp13/zhp1
; OTHER INFORMATION: of SEQ ID NO:6
; NAME/KEY: misc_feature
; LOCATION: (1)...(1731)
; OTHER INFORMATION: n = A,T,C or G
US-09-997-610-7
```

```
Query Match          70.4%; Score 57; DB 9; Length 1731;
Best Local Similarity 64.2%; Pred. No. 7.2e-10;
Matches 52; Conservative 9; Mismatches 20; Indels 0; Gaps 0;
```

```
OY 1 GGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCGTGACACAGGTTTA 60
    |||
DB 409 GGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCGTGACACAGGTTTA 468
OY 61 CCACATATACAGAGAATA 81
    |||
DB 469 CCNCATATACAGAGAATA 489
```

```
RESULT 5
US-09-960-352-1786
; Sequence 1786, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengping
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 1786
; LENGTH: 392
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 08-LIB34-023-Q1-E1-B7
US-09-960-352-1786
```

```
Query Match          40.0%; Score 32.4; DB 10; Length 392;
Best Local Similarity 64.9%; Pred. No. 0.069;
Matches 48; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
```

```
OY 1 GGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCGTGACACAGGTTTA 60
```

Db 155 GGCGCACGAGGACACCCAGGCGCTCCAGGTATAGAGGGCGCTCCAGGTATAGAGGATATA 214  
QY 61 CCACATATATACGAG 74  
Db 215 CCAGGTTTGCACAG 228

## RESULT 6

US-09-960-352-9137  
; Sequence 9137, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathalagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960,352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 9137  
; LENGTH: 415  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 39-LIB34-018-Q1-E1-B12  
US-09-960-352-9137

Query Match 40.0%; Score 32.4; DB 10; Length 415;  
Best Local Similarity 64.9%; Pred. No. 0.07;  
Matches 48; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1 GGACCTCCAGCACACCCAGGCGCTCCAGGTATAGAGGGCGCTCCAGGTATAGAGGATATA 60  
Db 165 GGCGCACGAGGACACCCAGGCGCTCCAGGTATAGAGGGCGCTCCAGGTATAGAGGATATA 224  
QY 61 CCACATATATACGAG 74  
Db 225 CCAGGTTTGCACAG 238

## RESULT 7

US-09-923-779-49  
; Sequence 49, Application US/09923779  
; Patent No. US20020076721A1  
; GENERAL INFORMATION:  
; APPLICANT: Pyle, Ruth A.  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.553  
; CURRENT APPLICATION NUMBER: US/09/923,779  
; CURRENT FILING DATE: 2001-08-06  
; NUMBER OF SEQ ID NOS: 155  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 49  
; LENGTH: 549  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 480, 498, 523, 539  
; OTHER INFORMATION: n = A,T,C or G  
US-09-923-779-49

Query Match 37.5%; Score 30.4; DB 10; Length 549;  
Best Local Similarity 60.5%; Pred. No. 0.35;  
Matches 49; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 GGACCTCCAGCACACCCAGGCGCTCCAGGTATAGAGGGCGCTCCAGGTATAGAGGATATA 60

Db 406 GGAATCTCCAGGATATACAGAGACCCCTGTGTAACCTGCGACACTGCTCTTACAGGCTT 465  
QY 61 CCACATATATACGAGGAATA 81  
Db 466 CCAGGACCTCCTGAGGCTATA 486

## RESULT 8

US-09-880-107-2178  
; Sequence 2178, Application US/09880107  
; Patent No. US20020142981A1  
; GENERAL INFORMATION:  
; APPLICANT: Horne, Darci T.  
; APPLICANT: Vockley, Joseph G.  
; APPLICANT: Scherf, Uwe  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
; FILE REFERENCE: 44921-5028-WO  
; CURRENT APPLICATION NUMBER: US/09/880,107  
; CURRENT FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: US 60/211,379  
; PRIOR FILING DATE: 2000-06-14  
; PRIOR APPLICATION NUMBER: US 60/237,054  
; PRIOR FILING DATE: 2000-10-02  
; NUMBER OF SEQ ID NOS: 3950  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2178  
; LENGTH: 3394  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 L22548  
US-09-880-107-2178

Query Match 37.5%; Score 30.4; DB 10; Length 3394;  
Best Local Similarity 71.4%; Pred. No. 0.6;  
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 GGACCTCCAGCACACCCAGGCGCTCCAGGTATAGAGGGCGCTCCAGGTATAGAGGATATA 56  
Db 670 GGCGCACGAGGACACCCAGGCGCTCCAGGTATAGAGGGCGCTCCAGGTATAGAGGATATA 725

## RESULT 9

US-09-764-868-48  
; Sequence 48, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT232  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 48  
; LENGTH: 2686  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (2505)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (2569)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (2644)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (2645)

```
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-48

Query Match          37.0%; Score 30; DB 9; Length 2686;
Best Local Similarity 61.5%; Pred. No. 0.77;
Matches 48; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 1 GGACCTCCAGCACACCCAGGCCCCAGAGAAAGTGGGCTCTCTGTCGACCAAGTTTA 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1305 GGAGCTCTGACACCGAGACCTATGTGAAGCGGCTGCACCTGCTGACCAAGTTT 1364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 61 CCACAATATACAGAGAA 78
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1365 CTGCACCAAGCTGACGGA 1382

RESULT 10
US-09-799-799-1
; Sequence 1, Application US/09799799
; Patent No. US20020132291A1
; GENERAL INFORMATION:
; APPLICANT: YE, Jane et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; FILE REFERENCE: CL001157
; CURRENT APPLICATION NUMBER: US/09/799,799
; CURRENT FILING DATE: 2001-03-07
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3380
; TYPE: DNA
; ORGANISM: Human
US-09-799-799-1

Query Match          37.0%; Score 30; DB 10; Length 3380;
Best Local Similarity 61.5%; Pred. No. 0.82;
Matches 48; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 1 GGACCTCCAGCACACCCAGGCCCCAGAGAAAGTGGGCTCTCTGTCGACCAAGTTTA 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1025 GGAGCTCTGACACCGAGACCTATGTGAAGCGGCTGCACCTGCTGACCAAGTTT 1084
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 61 CCACAATATACAGAGAA 78
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1085 CTGCACCAAGCTGACGGA 1102

RESULT 11
US-09-923-779-47
; Sequence 47, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 544..550
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-47
```

```
Query Match          36.8%; Score 29.8; DB 10; Length 561;
Best Local Similarity 60.5%; Pred. No. 0.56;
Matches 49; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy 1 GGACCTCCAGCACACCCAGGCCCCAGAGAAAGTGGGCTCTCTGTCGACCAAGTTTA 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 406 GGATCTCCAGATACCAAGAGACCCCTGCTGAACCTGGGCAAGCTGCTTCAGGCCCT 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 61 CCACAATATACAGAGAAATA 81
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 466 CCAGGACCTCTCTGCTATA 486

RESULT 12
US-09-923-779-51
; Sequence 51, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51
; LENGTH: 617
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 581..605
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-51

Query Match          36.8%; Score 29.8; DB 10; Length 617;
Best Local Similarity 60.5%; Pred. No. 0.57;
Matches 49; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy 1 GGACCTCCAGCACACCCAGGCCCCAGAGAAAGTGGGCTCTCTGTCGACCAAGTTTA 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 406 GGATCTCCAGATACCAAGAGACCCCTGCTGAACCTGGGCAAGCTGCTTCAGGCCCT 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 61 CCACAATATACAGAGAAATA 81
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 466 CCAGGACCTCTCTGCTATA 486

RESULT 13
US-09-923-779-66
; Sequence 66, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 648
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
US-09-923-779-66
```

```

; LOCATION: 642, 646
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-66

```

Query Match	36.8%;	Score 29.8;	DB 10;	Length 648;
Best Local Similarity	60.5%;	Pred. NO. 0.58;		
Matches 49;	Conservative 0;	Mismatches 32;	Indels 0;	Gaps 0;

QY 1 GGACTCCAGACACACCCCGAGGCCCCCAGAGAAGTGGGGCTCCTGTGTGACACAGTTTA 60  
 ||| ||||| ||| | ||||| ||| ||||| ||| |||  
 Db 406 GGATCTCCAGGATACCAAGAGGACCCCTGTGTGAACCTGGGCAAGTGTCTTCAGAGCCCT 465

Qy	61	CCACATATTCAGGAGAATA	81
Db	466	CCAGGACCTCCTGTCCTATA	486

RESULT 14  
US-09-923-779-53

```

; Sequence 53, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:

```

APPLICANT: pyle, Ruth A.  
APPLICANT: Xu, Jiangchu  
APPLICANT: Kalos, Michael  
APPLICANT: ...

```

;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
;; FILE REFERENCE: 210121.553

```

: CURRENT APPLICATION NUMBER: US/09/923,779  
 : CURRENT FILING DATE: 2001-08-06  
 : NUMBER OF SEQ ID NOS: 155

```

; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 53
; LENGTH: 653

```

```

; LENGTH: 653
;
; TYPE: DNA
; ORGANISM: Homo sapiens
;

```

```
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 517, 579, 581, 603, 649
```

OTHER INFORMATION: n = A,T,C or G  
US-09-923-779-53

Query Match	36.8%	Score 29.8	DB 10	Length 653
Best Local Similarity	60.5%	Pred. No. 0.58		
Matches 49: Conservative	0	Mismatches	32	Indels 0
				Gaps 0

QY 1 GGACCTCCAGCACACCCAGGCCCCCAGAGAAGTGGGGCTCCTGGTGCACACAGTTTA 60  
||| ||||| ||| ||||| ||| ||||| ||||| |||||  
Dp 406 GGATCTCCAGGATACCAAGACCCCTGGTGAACCTGGGCAAGCTGGTCCCTTCAGGGCCT 450

Qy	61	CCACATATACAGGAGAAATA	81
Db	466	CCAGGACCTCCTGGTCTATA	486

RESULT 15  
US-09-923-779-68

```

; Sequence 68, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:

```

APPLICANT: Pyle, Ruth A.  
APPLICANT: Xu, Jiangchun  
APPLICANT: Kalos, Michael

APPLICANT: kaioS, Michael D.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER  
FILE REFERENCE: 31031153

```

; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06

```

```

:  NUMBER OF SEQ ID NOS: 155
:
:  SOFTWARE: FastSeq for Windows Version 4.0
:
:  SEQ ID NO 68

```

LENGTH: 655  
TYPE: DNA

```

; ORGANISM: Homo sapiens
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 654
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-68

```

Query Match	36.8%;	Score 29.8;	DB 10;	length 655;
Best Local Similarity	60.5%;	Pred. No. 0.58;		
Matches 49;	Conservative 0;	Mismatches 32;	Indels 0;	Gaps 0;

QY 1 GGACTCCAGCACACCCAGGCCCCCAGAGAAGTGGGCTCTGGTGCACACAGTTTA 60  
 ||| ||||| | ||| | ||||| ||| ||||| | |||||  
 Db 406 GGATCTCCAGGATACCAAGAGACCCCTGGTGAACCTGGCAAGCTGCTTTCAGGCTT 455

Qy	61	CCACAATATACAGGAGGAATA	81
Db	466	CCAGGACCTCCTGCTCTATA	486

Search completed: February 19, 2003, 23:02:28  
Job time : 11.518 secs

```

: LOCATION: 642, 646
: OTHER INFORMATION: n = A,T,C or G
US-09-923-779-66

Query Match
Best Local Similarity 36.8%; Score 29.8; DB 10; Length 648;
Matches 49; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 GGACCTCCAGACACCCCGCCGAGAGAGTGGGCTCTGCTGTCACACGAGTTTA 60
DB 406 GGATCTCCAGAGATACCAAGAGACCCCTGGTGAACCTGGCAAGCTGCTCTTCAGGACCT 465
QY 61 CCACATATATACAGAGAAATA 81
DB 466 CCAGGACCTCTGCTGCTATA 486

RESULT 14
US-09-923-779-53
: Sequence 53, Application US/09923779
: Patent No. US20020076721A1
: GENERAL INFORMATION:
: APPLICANT: Pyle, Ruth A.
: APPLICANT: Xu, Jiangchun
: APPLICANT: Kalos, Michael D.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: FILE REFERENCE: 210121.553
: CURRENT APPLICATION NUMBER: US/09/923,779
: CURRENT FILING DATE: 2001-08-06
: NUMBER OF SEQ ID NOS: 155
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 53
: LENGTH: 653
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: 517, 579, 581, 603, 649
: OTHER INFORMATION: n = A,T,C or G
US-09-923-779-53

Query Match
Best Local Similarity 36.8%; Score 29.8; DB 10; Length 653;
Matches 49; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 GGACCTCCAGACACCCCGCCGAGAGAGTGGGCTCTGCTGTCACACGAGTTTA 60
DB 406 GGATCTCCAGAGATACCAAGAGACCCCTGGTGAACCTGGCAAGCTGCTCTTCAGGACCT 465
QY 61 CCACATATATACAGAGAAATA 81
DB 466 CCAGGACCTCTGCTGCTATA 486

RESULT 15
US-09-923-779-68
: Sequence 68, Application US/09923779
: Patent No. US20020076721A1
: GENERAL INFORMATION:
: APPLICANT: Pyle, Ruth A.
: APPLICANT: Xu, Jiangchun
: APPLICANT: Kalos, Michael D.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: FILE REFERENCE: 210121.553
: CURRENT APPLICATION NUMBER: US/09/923,779
: CURRENT FILING DATE: 2001-08-06
: NUMBER OF SEQ ID NOS: 155
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 68
: LENGTH: 655
: TYPE: DNA

```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 19, 2003, 14:53:22 ; Search time 1157.56 Seconds

(without alignments)  
2084.671 Million cell updates/sec

Title: US-09-997-610-2\_COPY\_1\_149  
Perfect score: 790  
Sequence: 1 IIVIPVLITAVIEHVEVAGP.....KEISKQSGIQTWVLLKA 149

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+P2n.model -DEV=xli  
-O=/cgn2\_1/USPTO.spool/US09997610/runat\_10022003\_160823\_23843/app-query.fasta.1.1635  
-DB=EST -QFMT=fastap -SUFFIX=first -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45  
-DOCLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pic -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09997610.ecgn.1.1181-ctunat.10022003\_160823\_23843 -KCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELExt=7

Database :  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query length	DB ID	Description
1	311	39.4	499	14	BM967732
2	228.5	28.9	526	12	BE757276
3	226.5	28.7	504	13	BE757275
4	181	22.9	663	13	B3068972
5	165.5	20.9	1092	17	CNS05408
6	158.5	20.1	512	13	B3494317
7	158	20.0	587	10	AV593019
8	157.5	19.9	689	10	BB318780
9	157.5	19.9	2462	11	AK018742
10	157	19.9	662	14	BQ419957
11	155.5	19.7	695	9	AV243093
12	154.5	19.6	589	14	B0193813
13	154.5	19.6	921	9	AA673154
14	152.5	19.3	468	10	AW988554
15	152.5	19.3	552	10	BE625509
16	152.5	19.3	633	10	BB625284
17	152.5	19.3	707	12	BB6872725
18	152.5	19.3	921	11	AK003138
19	151.5	19.2	460	9	A1332213
20	150.5	19.1	508	9	AA493090
21	150.5	19.1	885	9	A1323115
22	150	19.0	942	14	B0900894
23	149	18.9	589	13	B1760895
24	148.5	18.8	856	14	B0876848
25	147.5	18.7	1029	17	CNS022P5
26	146.5	18.5	607	12	BB691150
27	146	18.5	585	10	AW913280
28	145.5	18.4	580	12	BB691162
29	145.5	18.4	582	10	BE588278
30	145	18.4	430	10	AA453561
31	144.5	18.3	453	10	BE126709
32	144.5	18.3	639	13	AA060543
33	142.5	18.0	693	9	AA060543
34	141.5	17.9	890	14	B0737531
35	141	17.8	683	10	AA029321
36	140.5	17.8	529	12	BE074104
37	140	17.7	482	14	BQ260214
38	140	17.7	542	14	BQ449231
39	140	17.7	563	14	BQ285664
40	140	17.7	566	14	BQ258785
41	140	17.7	580	13	B1477496
42	140	17.7	587	14	BQ420027
43	140	17.7	597	14	BQ285786
44	140	17.7	598	14	BQ420183
45	140	17.7	626	14	BQ260132

## ALIGNMENTS

RESULT 1  
BM967732  
LOCUS BM967732 499 bp mRNA linear EST 20-MAR-2002  
DEFINITION LM24HM0134 Bos taurus LM-24-HW cDNA library Bos taurus cDNA clone  
ACCESION LM-24-HW-011-34 (5'), mRNA sequence.  
VERSION BM967732  
KEYWORDS EST.  
SOURCE  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE  
1 (bases 1 to 499)  
Yoon,D.H., Jang,Y.S., Kim,T.H., Park,E.W., Lee,H.K., Chung,E.R.,

TITLE Sun,S.S. and Cheong,I.C.  
 JOURNAL Gene Expression Profiling of the Bovine skeletal muscle  
 COMMENT Unpublished (2002)  
 Contact: Dr. Du-Hak Yoon  
 National Livestock Research Institute, RDA  
 564 Omoekchun-dong, Suwon, 441-350, Korea  
 Tel: 82 31 290 1593  
 Fax: 82 31 290 1792  
 Email: dhyoon@rda.go.kr  
 Insert Length: 499 Std Error: 0.00  
 Seq primer: CAGGAACAGCTATGAC

FEATURES  
 source Location/Qualifiers

1..499  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /clone\_lib="LM-24-HW-011-34 (5')"  
 /clone\_lib="Bos taurus LM-24-HW cDNA library"  
 /sex="six males mixed"  
 /tissue\_type="longissimus dorsi"  
 /cell\_type="myocyte"  
 /dev\_stage="24 months Old"  
 /lab\_host="XLI-BlueMR/strain"  
 /note="Organ: skeletal muscle; Vector: Uni-ZAPXR; Site\_1:  
 EcoRI; Site\_2: Xho I"

BASE COUNT 129 a 139 c 131 g 100 t  
 ORIGIN

#### Alignment Scores:

Pred. No.: 4,99e-23 Length: 499  
 Score: 311.00 Matches: 68  
 Percent Similarity: 63.49% Conservative: 12  
 Best Local Similarity: 53.97% Mismatches: 37  
 Query Match: 39.37% Indels: 9  
 DB: 14 Gaps: 3

US-09-997-610-2\_COPY\_1\_149 (1-149) x BM67732 (1-499)

QY 1 l l e v a l l i e p r o v a l l e u l e t h r a l a v a l l e g l u h i s -----ValGIuVal 17  
 Db 120 CTGGTCACATCATCTAATGACAAATGCTGCTCGAGAAAGAGGTACCGGACCT 179  
 QY 18 A l a g l y p r o p r o l a h i s p r o a r g p r o p r o g l u g l u a l g l y p r o p r o -----Gly 34  
 Db 180 CCGGGCCACGACGACACCCAGGGCTCCAGGTAAAGAGGCTCCAGGTATTAAGAGA 239  
 QY 35 A l a p r o g l y l e u p r o g l i n t y r t h r g l y l u i l e s e r g l u m e t h r l y s c y s p r o c y s p r o 54  
 Db 240 ATACAGGTTGGCAGAGTCGGCGGAGAACCCAGACCAAGTATAAATGCCCATGCCAC 299  
 QY 55 A s p l i e g l u a r g s e r a l a p h e t h r v a l l y s l e u s e r g l y l y s l e u p r o l e u p r o p h e l y s 74  
 Db 300 -----AGACAGCTCCGCTTCACCTGTAAGCTCAGTGGCCAGTTGCCCTTCCCTTCAAG 353  
 QY 75 P r o i l e -l l e p h e t h r g l y v a l l e u t y r a s n a l a g l n a r g a s p l e u l y s g l u a l a m e t g l 94  
 Db 354 C C T G G C C C T T C C A G A G G T C C T G T A C A A T G C C A G A A A G A C T T A C A G A G G A C A C T G G 413  
 QY 94 y v a l p h e a l a c y a r g v a l p r o g l y a s n t y r t y r s e r s e r p h e a s p v a l g l u e h i s h i 114  
 Db 414 G G T C T T C A C A T G C A G A G G T G C C A G A A T T A C A T T C C T T T T G A T G A T C G A T C C A T C A 473  
 QY 114 s c y l y s v a l a s n i l e 119  
 Db 474 C T G C A A G G T G A C T G T T 489

RESULT 2  
 BE757276 526 bp mRNA linear EST 25-APR-2001  
 LOCUS 211669 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.  
 DEFINITION BE757276  
 ACCESSION BE757276  
 VERSION BE757276.1 GI:10171268  
 KEYWORDS EST.

SOURCE  
 ORGANISM cow.  
 Bos taurus

REFERENCE  
 AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,  
 Casas,E., Wray,J.E., White,J., Cho,J., Fahrtenkrug,S.C., Bennett,  
 G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,  
 Pereira,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and  
 Keeler,J.W.  
 1 (bases 1 to 526)  
 Bovidae; Bovinae; Bos.

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA  
 JOURNAL libraries and construction of a gene index for cattle  
 MEDLINE Genome Res. 11 (4), 626-630 (2001)  
 COMMENT 21180013  
 Contact: Smith TPJ  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt-trimmed with phred  
 v0.980904.e. Vector identified by cross-match with the -minsore 18  
 and -minmatch 12 options.  
 PCR Primers  
 FORWARD: AGGAACAGCTATGACCAT  
 BACKWARD: GTTTCCTCAGTACAGC  
 Plate: 63 row: g column: 14  
 Seq primer: ATTGTAGTACACTATAG.

FEATURES  
 source Location/Qualifiers

1..526  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /clone\_lib="MARC 2BOV"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV SPORT6; Site\_1: NotI; Site\_2: SalI;  
 library made from pooled tissue from testis, thymus,  
 semitendinosus muscle, longissimus muscle, pancreas,  
 adrenal, and endometrium."

#### BASE COUNT

106 a 137 c 153 g 130 t  
 ORIGIN

US-09-997-610-2\_COPY\_1\_149 (1-149) x BE757276 (1-526)

QY 19 G l y p r o p r o -----A l a h i s p r o a r g p r o p r o g l u g l u a l g l y p r o p r o g l y a l a 35  
 Db 181 G G A C C T C C G G A C T A G T G C T T C C A G G C C A C C T G T T A C A G A C C A T A G A G T A T G 240  
 QY 36 P r o g l y l e u -----P r o g l i n t y r t h r g l y l u i l e s e r g l u m e t h r l y s c y s p r o 52  
 Db 241 C C A G G A C T A A C G G A G G A C C G G C T T C C T G A C T G A C G T T G A G A G ----- 285  
 QY 53 C y s p r o a s p l i e g l u a r g s e r a l a p h e t h r v a l l y s l e u s e r g l y l y s l e u p r o l e u p r o 72  
 Db 286 T G C C A C C C C T C C A G T C T G C T T T C C T T C C A A G C G A G A T G G C C T T T C C A G A C C C 345  
 QY 73 P h e l y s p r o i l e l l e p h e t h r g l y v a l l e u t y r a s n a l a g l n a r g a s p l e u l y s g l u a l a 92  
 Db 346 T C C C A G C C A T T G T C T T C C A G A A G T T C T G A C A C C A T C A G G C C A C T T G A C C C C G C C 405  
 QY 93 M e t g l y v a l p h e a l a c y a r g v a l p r o g l y a s n t y r t y r s e r s e r p h e a s p v a l g l u e u 112  
 Db 406 A C T G T G T T A C A G T G C A G C G C C C T G T G T A C C A C T T T G C C T T T G A C A T T G A C T T G 465

```

OY 113 H1SH1ScyLysValAsn1LeTrrpleuMetArgLysGln1le 126
Db 466 TTTCAGAGTGTCTCAGAGTGTCTTAATGCGAATGCGATC 507

RESULT 3
LOCUS BE757275 504 bp mRNA linear EST 25-APR-2001
DEFINITION 211668 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE757275
VERSION BE757275.1 GI:10171267
KEYWORDS EST.
SOURCE
ORGANISM Bos taurus
COV.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE
1 (bases 1 to 504)
Smith,T.P.L., Grose,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Petter,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keeler,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemall.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -m1nscore 18
and -m1match 12 options.
PCR primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCAGCTCAGCAGC
Plate: 63 row: G column: 13
Seq primer: ATTAGGTGACATATAG.
FEATURES
source
Location/Qualifiers
1..504
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 2BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site.1: NotI; Site.2: SalI;
library made from pooled tissue from testis, thymus,
semilendonsus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."
BASE COUNT 99 a 131 c 146 g 128 t
ORIGIN
Alignment Scores:
Pred. No.: 4.49e-14 Length: 504
Score: 226.50 Matches: 51
Percent Similarity: 55.86% Conservative: 11
Best Local Similarity: 45.95% Mismatches: 38
Query Match: 28.67% Indels: 11
DB: 12 Gaps: 3
US-09-997-610-2_COPY_1_149 (1-149) x BE757275 (1-504)
OY 19 glyProPro-----Alah1sProARgProProgluGluValGlyProProgluYala 35
Db 181 GGACCTCGGAGCAGTGTCTTCCAGGCGCCTGTTATACAGACCCATAGGTATG 240
OY 36 ProGlyLeu-----ProGlnTyrThrGlyGluLeuSerGluMetThrLysCysPro 52
Db 241 CCAGGACCAACAGGAGGAGCTGGCTTCTCTGACGTGTGAGACG----- 285
OY 53 CysProAspIleGluArgSerAlaPheThrValLysLeuSerGlyLysLeuProLeuPro 72

```

```

Db 286 TGCCACACCCCTGCTCAGTGTGCTTTCGTCACAGCTGAGTGGCTTTCCAGGACCC 345
OY 73 PheLysProIlellePheThrGlyValLeuTyrAsnAlaGlnArgAspLeuYsgLula 92
Db 346 TCCACACCCATGTTCTTCCAGGAAGTTCGTACACACCATCAGGCGCATTCGACCCGCC 405
OY 93 MetGluValPheAlaCysArgValProGlyAsnTyrTyrSerSerPheAspValGluLeu 112
Db 406 ACTGTTGTGTAGCTGACGACGCTCCCTGGTGTACCACTTGGCTTGCATGTGATTG 465
OY 113 H1SH1ScyLysValAsn1LeTrrpleuMetArg 123
Db 466 TTTCAGAGTGTCTCAGAGTGTCTTAATGCGG 498

RESULT 4
LOCUS BU068972 663 bp mRNA linear EST 11-DEC-2001
DEFINITION BU068972 NIBB Mochli normalized Xenopus tailbud library Xenopus
laevis cDNA clone XL051f08 5', mRNA sequence.
ACCESSION BU068972.1 GI:17497722
VERSION BU068972.1
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 663)
Kitayama,A., Terasaka,C., Mochli,M., Ueno,N., Shin-I,T. and Kohara
,Y.
Expressed genes in X. laevis embryo
Unpublished (2001)
Contact: Tadasu Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
FEATURES
source
Location/Qualifiers
1..663
/organism="Xenopus laevis"
/db_xref="taxon:9355"
/clone_lib="XL051f08"
/clone_lib="NIBB Mochli normalized Xenopus tailbud
library"
/tissue_type="whole embryo"
/dev_stage="stage 25"
/note="Vector: pBSRN3; Site.1: NotI; Site.2: EcoRI; cDNAs
were oligo-dT primed and directionally cloned. Staging
according to Nieuwkoop and Faber. Library is subcloned
and was constructed by N. Garrett and A.M. Zorn,
(Wellcome/CRC Institute)."
BASE COUNT 179 a 160 c 164 g 160 t
ORIGIN
Alignment Scores:
Pred. No.: 4.37e-09 Length: 663
Score: 181.00 Matches: 46
Percent Similarity: 42.55% Conservative: 14
Best Local Similarity: 32.62% Mismatches: 47
Query Match: 22.91% Indels: 34
DB: 13 Gaps: 6
US-09-997-610-2_COPY_1_149 (1-149) x BU068972 (1-663)
OY 18 AlaGlyProProAlaHisProARgProProgluGluValGlyProProgluYalaProGly 37
Db 106 TCAGGCCACCTGATGATGACAGACCTCA-----GGACACACAGTGAACCTGGC 156
OY 38 LeuPro-----GlnTyrThrGlyGluLeuSer 46

```

```

Db      157  CGTCCTGCAGAAATCGAGACTACGCCCCAGSTATGAGCAGATATTTCGCAAAATATGGA 216
Oy      47  GUmethrlyscysProCysProasp-----55
Db      217  GCCATTGATGAGTATMAAACGCTCATGATGACGAGAAAAGAAAATGAGAGA 276
Oy      56  -----11eGUarSerAlaphethrValLysLeuSerGlyLysLeuProLeuProPhe 73
Db      277  AACATTACAGAAATGCCAGCTTCACGCTGCAAGACGACGCCCTTCCTTATAGTGGGT 336
Oy      74  LysProLeuLeuPhePhePheGlyValLeuTyraPheAlaGlnArgAspLeuLysGluAlaMet 93
Db      337  GAACCTATTAAGTTTGAGAAACCTGTATACAGCGCAGAAATGGCTACAAATCCACTTACT 396
Oy      94  GLYValPheAlaCysArgValProGlyAsnTyTyrSerSerPheAspValGluLeuHis 113
Db      397  GGGGTGTACACTGTGAATTCAGGAATTTATTACTTTTCTTACCATGTC-----447
Oy      114  HisCysLys---ValAsnTleTyrLeuMetArgLysGlnLeuAlaAsnLysGluLys 132
Db      448  CATGTAAAGGTGCCAATGTTGATGAT-----GCGTTGTTCAGAAATATAGACCA 498
Oy      133  Ile 133
Db      499  GTA 501

RESULT 5
CNS0540B 1092 bp DNA linear GSS 26-JUL-2000
LOCUS Tetradon nigroviridis genome survey sequence T3 end of clone
DEFINITION 011D22 of library A from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION AL320996.1 GI:9535880
VERSION GSS: genome survey sequence.
KEYWORDS Tetradon nigroviridis.
SOURCE Tetradon nigroviridis.
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 1092)
AUTHORS Roest Crolius,H., Jallion,O., Dasilva,C., Ozouf-Costaz,C., Fizesmes,C.,
Bernot,A., Fizesmes,C., Wincker,P., Brotier,P., Quetier,F., Saurin,W.,
Saurin,W. and Weissenbach,J.
Estimate of human gene number provided by genome-wide analysis
using Tetradon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE 20296633
PUBMED 10835645
TITLE 2 (bases 1 to 1092)
AUTHORS Crolius,H.R., Jallion,O., Dasilva,C., Ozouf-Costaz,C., Fizesmes,C.,
Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W.,
Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
MEDLINE 20359837
PUBMED 10899143
TITLE 3 (bases 1 to 1092)
AUTHORS Genoscope.
JOURNAL Direct Submision
COMMENT Submitted (12-APR-2000)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetradon.
location/Qualifiers
FEATURES
source 1..1092
/organism="Tetradon nigroviridis"
/db_xref="taxon:99883"
/clone="011D22"
/clone_id="A"

```

---

```

BASE COUNT 224 a 301 c 308 g 245 t 14 others
ORIGIN
Alignment Scores:
Pred. No.: 3,9e-07 Length: 1092
Score: 165.50 Matches: 42
Percent Similarity: 45.80% Conservative: 18
Best Local Similarity: 32.06% Mismatches: 52
Query Match: 20.95% Indels: 19
DB: 17 Gaps: 4

US-09-997-610-2_COPY_1_149 (1-149) x CNS0540B (1-1092)
Oy      19  GLYProProAlaHisProArgProGluGluValGlyProGlyAla-----35
Db      528  GGGCCCCCTGGCCAGCCCTGGAGCCTGGAGCTGAGTGAAGCAATGGTGGCTGGTCT 587
Oy      36  ProGlyLeuProGlnTyrThrGlyGluLe-----Ser 46
Db      588  CCGGACCGCGCTGCTCTCTCGCGAAGTGTGTGGAGAAGGAATGGGTGGGTGAG 647
Oy      47  GUmethrlyscysProCysProaspPheGluArgSerAlaphethrValLysLeuSer 66
Db      648  GTTATGTCAAATCCCC-----ATGTCCTGCTTCACCTCCTCTGAC 692
Oy      67  GlyLysLeuProLeuProPheLysProLeuLeuPhePhePheGlyValLeuTyraPhe 86
Db      693  ACACCCATACCCAGCTGTGGCACCCCTATTAGTTTACCAAAATGTATACATGCTGAG 752
Oy      87  ArgAspLeuLysGluAlaMetGlyValPheAlaCysArgValProGlyAsnTyTyrSer 106
Db      753  AATCAGTAAACCCCGAGTCTGGCATCTTACCTGCAGGTTCTCGAGTCTACTCTTT 812
Oy      107  SerPheAspValGluLeuHis-----HisCysLysValAsnTleTyrLeuMetArgLys 124
Db      813  TCCATACAGATCCATGTCATGTCCTCATGCCCTGGTGGCTGTCAAAAATGGCCAG 872
Oy      125  GlnTleLeuAlaAsnLysGlnLulleSerLys 135
Db      873  CCTATTATGTTCACTTATGACGATACACAG 905

RESULT 6
Bj494317 512 bp mRNA linear EST 08-AUG-2002
LOCUS Bj494317 MF01FSA cDNA Oryzias latipes cDNA clone MF01FSA024E16 5',
DEFINITION mRNA sequence.
ACCESSION Bj494317.1 GI:22146243
VERSION Bj494317
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Alburninomorpha;
Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 512)
AUTHORS Kohara,Y., Shin-I,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
Medaka EST Project in Takeda's lab
JOURNAL unpublished (2001)
COMMENT Contact: Tadasu Shin-1
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
location/Qualifiers
FEATURES
source 1..512
/organism="Oryzias latipes"
/db_xref="taxon:8090"
/clone="MF01FSA024E16"

```

[illegible]

Single pass sequencing.  
This clone was obtained from a polyA-deleted cDNA library.

```

FEATURES
    source                1..587
                        /organism="Bos taurus"
                        /db_xref="taxon:9913"
                        /clone="E1CA005A05"
                        /clone_lib="Bos taurus cartilage fetus"
                        /tissue_type="cartilage"
                        /dev_stage="fetus"
                        /lab_host="DH10B"
                        /note="Vector: pZLL; Site_1: SalI; site_2: NotI; Poly A
                                was deleted from a NotI site"

BASE COUNT      114 a   161 c   170 g   138 t       4 others

ORIGIN
Alignment Scores:
Pred. No.:          1e-06           Length:         587
Score:             158.00           Matches:        40
Percent Similarity: 45.61%          Conservative:  12
Best Local Similarity: 35.09%       Mismatch:      44
Query Match:       20.00%           Indels:        18
DB:               10              Gaps:          4

US-09-997-610-2_COPY1_149 (1-149) x AV593019 (1-587)
OY      18  AAGAGTGTGTTAAATGAAGCGCAGCCAAAGGCCGTTGGNNAATGCCAATCAGGAGTA 37
          |||||  ||  |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      347  GCTGGGAGACCTGGCTCCAGAGCCCCCAGAGGCCCCAGAGGCCTCCAGAGCCAACAGC 288
OY      38  LeuProgin-----TYrThGlYluLleSerGlnet 48
          ||||||:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      287  CTGCCCGAGGACTTTGTAAAGCAGCCAAAGGCCGTTGGNNAATGCCAATCAGGAGTA 228
OY      49  ThrlyScySPrcCyPrGaSprlleglnArgerAlarPherThVallyLeuSerGilyls 68
          |||||  |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      227  ACAGGAAATCCCTGTG-----TCGCTTTTCACGTGTAATCTCTCCAAAGCT 183
OY      69  LeuProlaeurProRhyLySProllelleRhetRhgilyValLeuTyfAsnAlaGlnArgasp 88
          |||  |||  |||||  |||  ::|||  |||||  |||||  |||||  |||||  |||||
Db      182  TACCAGAGTATAAGTAGACATCATCCATCCATCGATAAGATTATTATATACAGCACAGCAT 123
OY      89  LeuLySGUAlaMetGlyValPheAlaCySarGvalAProGlyAsnTyTtySerSerPhe 108
          |||:|:|:|  |||:|:|:|  |||:|:|:|  |||||  |||||  |||||  |||||
Db      122  TATGACCCAGACAACTGGATCTTCACCTGTAAGATTCCAGAGCATATATACTTCTTAC 63
OY      109  AspValIGluLeuHSHiscyslys---ValAsnleIetPleu 121
          :|:|:|  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      62  CACATA-----CACGTGAAGAAGGACCCCATGCTTGGA 30

RESULT 8
LOCUS      BB318780                      689 bp     mRNA       linear     EST 24-OCT-2001
DEFINITION BB318780 RIKEN full-length enriched, adult male corpora
quadrigenima Mus musculus cDNA clone B2J3077C02 3 , mRNA sequence.
ACCESSION  BB318780
VERSION    BB318780.2 GI:16403219
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 689)
AUTHORS   Arakawa,T., Carlincl,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.A., Kouda
            , Hiramoto,K., Horl,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            , M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            , D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
COMMENT   Unpublished (2001)
On Jul11, 2000 this sequence version replaced gi:9025815.

```

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)

Wag1, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura  
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1737-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Salto, T., Kiyosawa, H., Yamanka, I., Alizawa  
K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and  
Hayashizaki, Y.

Computational Analysis of Full-length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp/) for  
further details.

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in RIKEN  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

#### FEATURES

##### source

Location/Qualifiers

1..689

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="B230377C02"

/clone\_lib="RIKEN full-length enriched, adult male corpora  
quadrigemina"

/sex="male"

/tissue\_type="corpora quadrigemina"

/dev\_stage="adult"

/lab\_host="DH10B"

/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'

GAGAGAGAGATCCACAGAGCTCTTTTCTTTTCTTTTCTTTVN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 20.0 and subtraction to Rot = 45.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGAGAGATCTCGATTAATTAATTAATTAATCCCCCCCC  
3']. cDNA was cleaved with XhoI and BamHI. Vector: a  
modified pluescript KS(+) after bulk excision from lambda

FLC I."

BASE COUNT 169 a 189 c 182 g 147 t 2 others

##### ORIGIN

##### Alignment Scores:

Pred. No.: 1,42e-06 Length: 689  
Score: 157.50 Matches: 40  
Percent Similarity: 43.20% Conservative: 14

Best Local Similarity: 32.00% Mismatches: 45  
Query Match: 19.94% Indels: 25  
DB: 10 Gaps: 4

US-09-997-610-2\_COPY\_1\_149 (1-149) x BB318780 (1-689)

QY 19 GlyProProAlaHisProArProProGluGluValGlyProProGlyAlaProGlyLeu 38

DB 89 GGGACGCTGGCTTCCTCGACCCCGAGGTCCTCGAGACCCCGAGGCTTCGAGTGTG 148

QY 39 -----ProGlnTrpThrGlu----- 44

DB 149 ATGCTTACACATACACCCAGGAGAGATTCGCCAGATATGAGCACTAGCATGATGCGG 208

QY 45 -----IleSerGluMetTrpTrpCysProCysProAspIleGlu 57

DB 209 GTGAACCTCCGCGATGCTATGCGGCAAAAGGCAACGAGGAGCCAGCTATGAG 268

QY 58 ArgSerAlaPheTrpValLysLeuSerGlyLysLeuProLeuProPheLysProIleIle 77

DB 269 ATGCTTGGCTTACTGCGAGCTGACTGATACCTTCCACCGGTGGGGCCAGTGAAG 328

QY 78 PheTrpGlyValLeuTrpValAsnAlaGlnArgAspLeuLysGluAlaMetGlyValPheAla 97

DB 329 TTTCACAGAGCTGCTTACACAGCGACAGACAGACTACATCCGACAGACGATTTGACC 388

QY 98 CysArgValProGlyAsnTrpTrpSerSerPheAspValGluLeuHisLysCysLys--- 116

DB 389 TGTGAAGTCCCGGGGTGTCTATGACTTTGCTTATCATGTT-----CACTGCACAGGGA 429

QY 117 ValAsnIleTrpLeu 121

DB 440 GGCACGATATGGGT 454

RESULT 9

AK018742

LOCUS AK018742 2462 bp mRNA linear HTC 19-JAN-2002

DEFINITION Mus musculus adult male kidney cDNA, RIKEN full-length enriched

library, clone:0610011L15;procollagen, type VIII, alpha 1, full

insert sequence.

ACCESSION AK018742

VERSION AK018742.1 GI:12858612

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (strain:C57BL/6J) adult male kidney cDNA to mRNA,

clone\_lib:RIKEN full-length enriched mouse cDNA library

clone:0610011L15.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

JOURNAL 99279253

PUBMED 10349636

REFERENCE 2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL 20499374

PUBMED 11042159

REFERENCE 3

Shibata, K., Itoh, M., Alizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,

Kono, H., Akiyama, J., Nishi, K., Kikunishi, T., Tashiro, H., Itoh, M.,

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,

Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer







of the final driver population. b) a pool of about 29,000 clones from subtracted libraries CA0 and CA1 corresponding to plates R-CA0-AMV through R-CA0-AXS, R-CA0-AXZ through R-CA0-BAZ, R-CA0-BFE through R-CA0-BH, R-CA0-BJS, R-CA0-BKE, R-CA0-BKG-H, R-CA0-BKJ-K, R-CA0-BKP through R-CA0-BKS, R-CA0-BKV-V, R-CA0-BLY through R-CA0-BMA, R-CA0-BMC through R-CA0-BME, R-CA0-BNS, R-CA0-BOB through R-CA0-BOJ, R-CA0-BPA through R-CA0-BPG, R-CA1-BBA through R-CA1-BDA, R-CA1-BH2 through R-CA1-BTP, R-CA1-BTR, R-CA1-BTJ through R-CA1-BKB, R-CA1-BKO, R-CA1-BKF, R-CA1-BKI, R-CA1-BKT, R-CA1-BLF, R-CA1-BLH through R-CA1-BLN, R-CA1-BLS, R-CA1-BLV-V, R-CA1-BM, and R-CA1-BLE. The resulting pool represented 20% of the final driver population. c) a pool of about 15,000 clones from non-normalized libraries CS0s, CT0s, CU0s, and CX0s and normalized libraries CS0s, CT0s, CU0s, and CX0s corresponding to plates R-CT0s-CAV, R-CU0s-CBP through R-CT0s-CAM through R-CT0s-CAV, R-CU0s-CBP through R-CU0s-CCA, R-CM0s-CCB through R-CM0s-CCM, R-CX0s-CCN through R-CX0s-CCX, R-CS0-BSD, R-CS0-BTD through R-CS0-BTV, R-CS0-BVM, R-CT0-BTW through R-CT0-BUP, R-CT0-BVN, R-CU0-BUQ through R-CU0-BXO, R-CM0-BVY through R-CM0-BMP, R-CM0-BXN through R-CM0-BXO, R-CX0-BMO through R-CX0-BXM. The resulting pool represented 5% of the final driver population. d) a pool of about 5,000 clones (1,000 from non-normalized eye library CVO and 4,000 from normalized eye library CV1) corresponding to plates R-CV1-BSC, R-CV1-BSE through R-CV0-BRR, R-CV1-BRS through R-CV1-BVO, R-CV1-BSE through R-CV1-BRC, and R-CV1-BVO through R-CV1-BVO. This pool represented about 5% of the final driver population. e) A pool of about 10,000 clones from subtracted library BS2, BVO and BVOP (7-9.5 kb cDNA library fraction from rat whole embryo), and BXO (0.5-7kb cDNA library fraction from rat whole embryo) corresponding to plates R-BS2-BDB through R-BS2-BFB, R-BVO-ANK through R-BVO-ANK, R-BVOP-AOI through R-BVOP-AOX, and R-BXO-AOY through R-BXO-ASH. The resulting pool represented 5% of the final driver population. f) a pool of about 7,000 clones from the seven non-normalized libraries that make up the tracer including CY0, CZ0, DA0, DB0, DC0, DD0, and DE0 corresponding to plates R-CY0-BXP through R-CY0-BXZ, R-CZ0-BYA through R-CZ0-BYI, R-CZ0-BZB-C, R-DA0-BXJ through R-DA0-BYP, R-DA0-BZD through R-DA0-BZH, R-DB0-BXQ through R-DB0-BZA, R-DC0-BZI through R-DC0-BZQ, R-DD0-CAY through R-DD0-CBA, R-DD0-BZR through R-DD0-CAA, R-DD0-CBB-C, and R-DE0-CAB through R-DE0-CAL. The resulting pool represented about 10% of the final driver population. g) a pool of about 2,000 clones from the pool of normalized libraries, CN0, that makes up the tracer. The corresponding plates are R-CN0-BKV through R-CN0-BLD, R-CN0-BLG, R-CN0-BLP through R-CN0-BMF through R-CN0-BML, R-CN0-BLM-X, R-CN0-BMB, and R-CN0-BMF through R-CN0-BML. This pool represented 5% of the final driver population. h) a pool of the 28 most abundant clones in the CN0 pool corresponding to the following addresses: bkx-a-09-0-UI, bkx-b-09-0-UI, bkx-b-11-0-UI, bkx-b-10-0-UI, bkx-d-01-0-UI, bkx-d-06-0-UI, bkx-g-08-0-UI, bkx-h-12-0-UI, bkx-a-05-0-UI, bkx-a-06-0-UI, bkx-a-11-0-UI, bkz-c-06-0-UI, bkz-a-05-0-UI, bkz-g-08-0-UI, bla-a-01-0-UI, bkz-c-09-0-UI, bkz-d-10-0-UI, bla-g-07-0-UI, bla-g-12-0-UI, bla-a-02-0-UI, bla-f-04-0-UI, bla-g-07-0-UI, bla-g-12-0-UI, bla-a-12-0-UI, bid-f-02-0-UI, bla-a-11-0-UI, bla-e-95-0-UI, bid-l-08-0-UI, bid-f-02-0-UI, bid-h-04-0-UI, bla-a-05-0-UI, blt-f-08-0-UI. This pool represented 5% of the final driver population. i) One abundant CN0 clone (corresponding to the address bkz-a-11-0-UI) was digested with Not I and Eco RI and the resulting insert was gel purified. This purified insert was added directly to the driver so that it represented 5% of the final driver population.

TAG LIB-UI-R-CN1  
TAG TISSUE=fundus  
TAG\_SEO=FTCGG"

BASE COUNT 119 a 149 c 170 g 151 t

## ORIGIN

## Alignment Scores:

Pred. No.:	2,366-06	Length:	589
Score:	154.50	Matches:	42
Percent Similarity:	41.04%	Conservative:	13
Best Local Similarity:	31.34%	Mismatches:	40
Query Match:	19,568	Indels:	39
DB:	14	Gaps:	6

US-09-997-610-2\_COPY\_1\_149 (1-149) x B0193813 (1-589)

QY	14	HISVALGUVLALAGLPPROPALHISPROARGPROGUGLVALGLYPROPTO	33
DB	587	CACGAGGCGCTCTCTGGGACCAAGGCCCTCCAGGCCCCCA	537
QY	34	GLYALA---PROGLYLEUPROGLNLYRTHGLGLULLESGLMETHTLYSCYSPRO	52
DB	536	GCTGTGATGCCCTTACACCATCACCCAGGAGATAT	498
QY	53	CYSPROAPLIE-----	56
DB	497	CTACCAAGATATGGACATAGATGAGTGACCTCCACATGCTATGCGGCAAA	438
QY	57	-----GLUARGSERALAPHETVALLYLSEUSERYLYS	68
DB	437	AAGGCAGACAGGAGGCGCTATGAGATGCCCGCTTACTGCGAGCTACTACT	378
QY	69	LEUPROLEUPROPHLYPROILEILEPHEHTHGLYVALLEUPLYRASNALAGINARGASP	88
DB	377	TTCACACCGCTGGGCGCCCAAGTAGATTGACAGAGTCTTATACATGAGACAGAC	318
QY	89	LEULYGLIUALMETCLYVALPHEALACYSARGVALPROGLYSNTLYRTERSERPHE	108
DB	317	TACACCCGACGAGGAGGTCTTACCTGCCCAACTCCGGGTCTACTACTTCTTAC	258
QY	109	ASPVALGLULHSHISCYLSYS---VALASNILETRPLEU	121
DB	257	CANGTT-----CAGTCAGAGGAGGACACTCTGGTT	225

RESULT 13  
AA673154  
LOCUS  
DEFINITION  
IMAGE:820668.5 similar to gb:U37222 Mus musculus 30kDa adipocyte complement-related protein Acrp30 (MOUSE);, mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

house mouse.  
Mus musculus.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 921)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
The WashU-HMNI Mouse EST project  
Unpublished (1996)  
Contact: Maria M/Mouse EST Project  
WashU-HMNI Mouse EST Project  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:488948  
Seq primer: -28n13 rev2 ET from Amerisham  
High quality sequence stop: 499.

FEATURES  
source

Location/Qualifiers  
1. .921  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_image="820668"  
/clone\_11b="Soares\_mammary\_gland\_NBMG"  
/sex="male"  
/tissue\_type="mammary gland"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5', TGTTACCAATCTGAAGTGGAGCGCCCGAATGTTTTTTTTTTTTTTTTTTT 3'); double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 235 a 230 c 237 g 219 t  
ORIGIN

Alignment Scores:  
Pred. No.: 4.47e-06 Length: 921  
Score: 154.50 Matches: 39  
Percent Similarity: 49.58% Conservative: 20  
Best Local Similarity: 32.77% Mismatches: 53  
Query Match: 19.56% Indels: 7  
DB: Gaps: 2

US-09-997-610-2\_COPY\_1\_149 (1-149) x AA673154 (1-921)

Oy 31 GtPProGlyAlaProGlyLeuProGlyTThrGlyGluIleSerGluMetThrLys 50  
||||| ||| ||||| ||| ||||| |||  
Db 307 GGGCCAGGGGCTCCCGGACCCCTGGCAGAAAGAGAGCGCTGAGAGCGCTTAT 366  
Oy 51 CysProGlyProAspIleGluArgSerAlaPheThrValLysLeuSerGlyLysLeuPro 70  
::: ||||| ||||| ||||| |||  
Db 367 -----GTGATCGCTCAGCGCTCAGTGGGCTGGAGACCCGCTGACT 411  
Oy 71 LeuProPheLysProIleIlePheThrGlyValLeuTyrAsnAlaGlnArgSpleuLys 90  
::: ||| ||||| ||||| ||||| |||  
Db 412 GTTCCCAATGTACCATTCGCTTTACTAAGATCTTCTACACCAACAGATCATTAATGAC 471  
Oy 91 GluAlaMetGlyValPheAlaCysArgValProGlyAsnTyrTyrSerSerPheAspVal 110  
::: ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 472 GGCAGCAGTGGCAAGTGTACTGCAACATTCGGGACTCTACTACTCTCTTACCCATC 531  
Oy 111 GluLeuHis-----HisCysLysValAsnIleTyrPleuMetArgLysGlnIleLeuAla 128  
::: ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 532 ACGGTACATGAAAGATGTGAAGTGAAGCTTTCACAGAGAGCAAGCCGTTCTCTC 591  
Oy 129 AsnLysGluGluIleSerLysGlnGlnIleGlnGluValThrTyrPalleuLeu 147  
::: ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 592 ACCTACGACCACTATCAGAGAAAGATGTGGACAGGCGCTGTGCTGCTCTC 648

RESULT 14  
LOCUS AW988554 468 bp mRNA linear EST 02-JUN-2000  
DEFINITION u008h07.y1 Soares\_mammary\_gland\_NMLMG Mus musculus cDNA clone  
IMAGE:1531069.5" similar to gb:U37222 Mus musculus 30kDa adipocyte  
complement-related protein Acip30 (MOUSE);, mRNA sequence.  
ACCESSION AW988554  
VERSION AW988554.1 GI:8183731  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 468)

AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
MGI:947169  
Seq primer: -40RP from Gibco  
High quality sequence stop: 457.

FEATURES  
source

Location/Qualifiers  
1. .468  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_image="1531069"  
/clone\_11b="Soares\_mammary\_gland\_NMLMG"  
/sex="female (lactating)"  
/tissue\_type="mammary gland"  
/lab\_host="DH10B"  
/note="vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 114 a 117 c 131 g 106 t  
ORIGIN

Alignment Scores:  
Pred. No.: 2.77e-06 Length: 468  
Score: 152.50 Matches: 39  
Percent Similarity: 49.58% Conservative: 20  
Best Local Similarity: 32.77% Mismatches: 53  
Query Match: 19.30% Indels: 7  
DB: Gaps: 2

US-09-997-610-2\_COPY\_1\_149 (1-149) x AW988554 (1-468)

Oy 31 GtPProGlyAlaProGlyLeuProGlyTThrGlyGluIleSerGluMetThrLys 50  
||||| ||| ||||| ||| ||||| |||  
Db 61 GGGCCAGGGGCTCCCGGACCCCTGGCAGAAAGAGAGCGCTGAGAGCGCTTAT 120  
Oy 51 CysProGlyProAspIleGluArgSerAlaPheThrValLysLeuSerGlyLysLeuPro 70  
::: ||||| ||||| ||||| |||  
Db 121 -----GTGATCGCTCAGCGCTCAGTGGGCTGGAGACCCGCTGACT 165  
Oy 71 LeuProPheLysProIleIlePheThrGlyValLeuTyrAsnAlaGlnArgSpleuLys 90  
::: ||| ||||| ||||| ||||| |||  
Db 166 GTTCCCAATGTACCATTCGCTTTACTAAGATCTTCTACACCAACAGATCATTAATGAC 225  
Oy 91 GluAlaMetGlyValPheAlaCysArgValProGlyAsnTyrTyrSerSerPheAspVal 110  
::: ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 226 GGCAGCAGTGGCAAGTGTACTGCAACATTCGGGACTCTACTACTCTCTTACCCATC 285  
Oy 111 GluLeuHis-----HisCysLysValAsnIleTyrPleuMetArgLysGlnIleLeuAla 128  
::: ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 286 ACGGTACATGAAAGATGTGAAGTGAAGCTTTCACAGAGAGCAAGCCGTTCTCTC 345  
Oy 129 AsnLysGluGluIleSerLysGlnGlnIleGlnGluValThrTyrPalleuLeu 147  
::: ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 346 ACCTACGACCACTATCAGAGAAAGATGTGGACAGGCGCTGTGCTGCTCTC 402

RESULT 15  
LOCUS BE625509 552 bp mRNA linear EST 24-AUG-2000  
DEFINITION u019e06.y1 Soares\_mammary\_gland\_NMLMG Mus musculus cDNA clone  
IMAGE:3372418.5" similar to gb:U37222 Mus musculus 30kDa adipocyte  
complement-related protein Acip30 (MOUSE);, mRNA sequence.  
ACCESSION BE625509  
VERSION BE625509.1 GI:9905925



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame-plus-p2n model

Run on: February 19, 2003, 14:57:18 ; Search time 627.004 Seconds

(without alignments)  
2088.706 Million cell updates/sec

Title: US-09-997-610-2\_COPY\_1\_45  
Perfect score: 240  
Sequence: 1 IVVIPLITAVIEHVEVAGP.....PPEVGGPGAGLPGYTGEL 45

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame-p2n model -DEV=x1h  
-Q/cgn2\_1/USPIO.spool/US09997610/runal\_10022003\_160823\_23819/app\_query.fasta.1.1635  
-DB=genbml -QMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45  
-DOCLALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09997610.qcgn.1.1.3965 -runat\_10022003\_160823\_23819 -NCPU=6 -ICPU=3  
-NO.XLPYX -NO.MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
Genbml:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_lnu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pin:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_sy:\*  
39: em\_hcgo\_hum:\*  
40: em\_hcgo\_mus:\*  
41: em\_hcgo\_other:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	197	82.1	145880	9 HS302D9	282198 Human DNA s
2	89	37.1	66669	3 AF017777	AF017777 Drosophill
3	89	37.1	87089	2 AC013980	AC013980 Drosophill
4	89	37.1	171548	3 AC092494	AC092494 Drosophill
5	89	37.1	173508	3 AC011251	AC011251 Drosophill
6	89	37.1	292911	3 AE003568	AE003568 Drosophill
7	87.5	36.5	68661	3 AC024805	AC024805 Caenorhab
8	87.5	36.5	278007	2 AC006799	AC006799 Caenorhab
9	86	35.8	4428	6 AX146422	AX146422 Sequence
10	86	35.8	4428	6 AX146424	AX146424 Sequence
11	85.5	35.6	5358	5 AY052763	AY052763 Xenopus 1
12	85	35.4	256417	2 AC079422	AC079422 Mus muscu
13	84	35.0	4428	6 AX146430	AX146430 Sequence
14	83	34.6	186	6 AR014097	AR014097 Sequence
15	83	34.6	186	6 AR117066	AR117066 Sequence
16	83	34.6	186	6 I06482	I06482 Sequence 19
17	83	34.6	186	6 I20524	I20524 Sequence 62
18	83	34.6	287	3 O05420010	AJ420010 Osterlagi
19	83	34.6	3612	10 MMCOLA4	X067771 Mouse mRNA
20	83	34.6	6512	6 AX306181	AX306181 Sequence
21	83	34.6	6512	10 MMCOLA4A	J04694 Mus muscu
22	83	34.6	36724	1 SC2G18	AL390188 Streptomy
23	83	34.6	43735	3 CBRG02P14	AC084481 Caenorhab
24	82.5	34.4	437	10 MMA379COL	X91012 M.musculus
25	82.5	34.4	1009	9 BC004412	BC004412 Homo sapi
26	82.5	34.4	1521	10 AF237721	AF237721 Mus muscu
27	82.5	34.4	1605	9 BC030945	BC030945 Mus muscu
28	82.5	34.4	2091	9 BC007574	BC007574 Homo sapi
29	82.5	34.4	2388	10 AF345718	AF345718 Mus muscu
30	82.5	34.4	2437	9 BC011705	BC011705 Homo sapi
31	82.5	34.4	2480	9 HUMCOL3IX	L41162 Homo saplen
32	82.5	34.4	2543	6 AR014074	AR014074 Sequence
33	82.5	34.4	2543	6 AR11836	AR11836 Sequence
34	82.5	34.4	6246	9 HSCOLPA3S2	AF026802 Homo sapi
35	82.5	34.4	69252	2 ACT01519	AC101519 Mus muscu
36	82.5	34.4	160241	9 HS885L7	AL035669 Human DNA
37	82.5	34.4	226854	10 AL669926	AL669926 Macaca mu
38	82	34.2	426	9 AF230926	AF230926 Sequence
39	82	34.2	994	6 AX067354	AX067354 Sequence
40	82	34.2	1485	10 MSC3A1A	M18933 Mouse alpha
41	82	34.2	1881	6 A60690	A60690 Sequence 20
42	82	34.2	1881	6 AR123971	AR123971 Sequence
43	82	34.2	2718	8 BC029697	BC029697 Mus muscu
44	82	34.2	3234	9 HSC3A1R	X15332 Human COL3A
45	82	34.2	3902	9 BC028178	BC028178 Homo sapi

RESULT 1

## ALIGNMENTS

HS302D9 145880 bp DNA linear PRI 12-DEC-1999  
LOCUS Human DNA sequence from clone RP1-302D9 on chromosome 22 Contains  
DEFINITION GSSs, complete sequence.  
ACCESSION Z82198  
VERSION Z82198.2 GI:6572207  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 145880)  
REFERENCE  
AUTHORS Bridgeman, A.  
JOURNAL Direct Submission  
Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Dec 13, 1999 this sequence version replaced gi:3164067.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMEP; Information  
on the WORMEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 22, constructed by the Sanger Centre Chromosome 22  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr22  
RP1-302D9 is from the library RPC1-1 constructed at the Roswell  
Park Cancer Institute by the group of Pieter de Jong. For further  
details see http://bacpac.med.buffalo.edu/  
VECTOR: pCYPAC2  
This sequence is the entire insert of clone RP1-302D9 the true left  
end of clone CTA-282F2 is at 69682 in this sequence. The true right  
end of clone CTA-415G2 is at 55167 in this sequence.  
FEATURES  
SOURCE location/Qualifiers  
1..145880  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="22"  
/clone="RP1-302D9"  
/clone\_1lb="RPC1-1"  
188..245  
repeat\_region /note="MER3 repeat: matches 144..209 of consensus"  
246..571  
repeat\_region /note="AluX repeat: matches 1..312 of consensus"  
572..759  
repeat\_region /note="MER3 repeat: matches 1..144 of consensus"  
783..933  
repeat\_region /note="MER5A repeat: matches 26..187 of consensus"  
1033..1336  
repeat\_region /note="AluSp repeat: matches 1..229 of consensus"  
1450..1583  
repeat\_region /note="MIR repeat: matches 24..160 of consensus"  
1687..1752  
repeat\_region /note="L2 repeat: matches 2593..2661 of consensus"  
2350..2660  
repeat\_region /note="AluSc repeat: matches 3..309 of consensus"  
2684..2981  
repeat\_region /note="AluSq repeat: matches 2..300 of consensus"  
3323..3343  
repeat\_region /note="MIR1E repeat: matches 116..136 of consensus"

repeat\_region 3344..3652  
/note="AluY repeat: matches 1..309 of consensus"  
repeat\_region 3653..3928  
/note="MIR1E repeat: matches 136..359 of consensus"  
repeat\_region 3929..4278  
/note="PHE1B repeat: matches 3..364 of consensus"  
4279..4485  
repeat\_region /note="MIR1E repeat: matches 359..568 of consensus"  
5073..5176  
repeat\_region /note="52 copies 2 mer ct 78 conserved"  
5181..5491  
repeat\_region /note="AluJb repeat: matches 1..311 of consensus"  
6369..6485  
repeat\_region /note="L2 repeat: matches 2579..2705 of consensus"  
6647..6685  
repeat\_region /note="MAD1 repeat: matches 1..23 of consensus"  
6686..6987  
repeat\_region /note="AluX repeat: matches 1..302 of consensus"  
6988..7036  
repeat\_region /note="MAD1 repeat: matches 23..77 of consensus"  
7482..7754  
repeat\_region /note="AluJb repeat: matches 9..290 of consensus"  
7775..8060  
repeat\_region /note="AluJo repeat: matches 1..295 of consensus"  
8414..8551  
repeat\_region /note="L2 repeat: matches 2553..2706 of consensus"  
8914..9030  
repeat\_region /note="MIR repeat: matches 147..262 of consensus"  
9110..9280  
repeat\_region /note="MIR repeat: matches 91..262 of consensus"  
9283..9412  
repeat\_region /note="MIR repeat: matches 15..144 of consensus"  
9521..9679  
repeat\_region /note="PAM repeat: matches 3..161 of consensus"  
9820..10225  
repeat\_region /note="MSTR repeat: matches 2..425 of consensus"  
complement(10179..10678)  
/note="match: GSS: Em:B56592"  
complement(10204..10728)  
/note="match: GSS: Em:A0701486"  
complement(10249..10706)  
/note="match: GSS: Em:A0225495"  
10312..10383  
repeat\_region /note="MIR repeat: matches 79..150 of consensus"  
10718..11310  
repeat\_region /note="match: GSS: Em:B14024"  
10784..11201  
repeat\_region /note="match: GSS: Em:B43656"  
11838..11946  
repeat\_region /note="MIR repeat: matches 20..137 of consensus"  
12174..12445  
repeat\_region /note="L2 repeat: matches 1988..2275 of consensus"  
12444..12642  
repeat\_region /note="MIR repeat: matches 63..241 of consensus"  
13017..13369  
repeat\_region /note="match: STS: Em:G49301"  
13331..13397  
repeat\_region /note="MIR repeat: matches 174..244 of consensus"  
13398..13698  
repeat\_region /note="AluSp repeat: matches 1..302 of consensus"  
13699..13810  
repeat\_region /note="MIR repeat: matches 76..174 of consensus"  
13806..13919  
repeat\_region /note="MIR repeat: matches 77..189 of consensus"  
13945..14060  
repeat\_region /note="MIR repeat: matches 24..142 of consensus"  
14061..14367  
repeat\_region /note="AluY repeat: matches 1..301 of consensus"  
14368..14452  
repeat\_region /note="MIR repeat: matches 141..225 of consensus"  
14589..14679  
repeat\_region /note="MIR repeat: matches 173..262 of consensus"  
14597..15201  
misc\_feature

```
misc_feature      14616..15060
                  /note="match: GSS: Em: A0553482"
                  /note="match: GSS: Em: A0370601"
repeat_region     14868..15040
                  /note="MIR repeat: matches 49. .233 of consensus"
repeat_region     15071..15188
                  /note="L2 repeat: matches 2112. .2239 of consensus"
repeat_region     15304..15399
                  /note="MLTIB repeat: matches 1. .99 of consensus"
repeat_region     15490..15662
                  /note="Alusg1 repeat: matches 2. .114 of consensus"
repeat_region     15699..15727
                  /note="MLTIB repeat: matches 119. .178 of consensus"
repeat_region     15728..16027
                  /note="AlusC repeat: matches 1. .299 of consensus"
repeat_region     16028..16245
                  /note="MLTIB repeat: matches 178. .390 of consensus"
repeat_region     16346..16854
                  /note="AluY repeat: matches 1. .300 of consensus"
repeat_region     18296..18323
                  /note="MSTR repeat: matches 2. .29 of consensus"
repeat_region     18324..18392
                  /note="MER66-internal repeat: matches 4919. .4993 of
consensus"
repeat_region     18393..18712
                  /note="Aluub repeat: matches 1. .311 of consensus"
repeat_region     18713..19133
                  /note="MER66-internal repeat: matches 4548. .4919 of
consensus"
misc_feature      complement(18872..19230)
                  /note="match: GSS: Em: A0005063"
                  19251..19719
misc_feature      /note="match: GSS: Em: B14179"
                  19537..20290
repeat_region     /note="HERFH21 repeat: matches 4657. .5784 of consensus"
                  20317..20382
repeat_region     /note="33 copies 2 mer ta 68 conserved"
                  20513..20666
repeat_region     /note="77 copies 2 mer tt 70 conserved"
                  20682..21008
repeat_region     /note="Alusg1 repeat: matches 1. .306 of consensus"
                  21239..21553
repeat_region     /note="HUEKS-P3 repeat: matches 4410. .4713 of consensus"
                  21882..22254
repeat_region     /note="THEIB repeat: matches 1. .364 of consensus"
                  22302..22537
repeat_region     /note="MER66-internal repeat: matches 2186. .2417 of
consensus"
                  22538..22850
repeat_region     /note="Alusp repeat: matches 1. .313 of consensus"
                  22851..23801
repeat_region     /note="MER66-internal repeat: matches 1210. .2186 of
consensus"
                  23905..23989
repeat_region     /note="MER66-internal repeat: matches 3017. .3102 of
```

```
Alignment Scores:
Pred. No.:      3.12e-07      Length:      145880
Score:          197.00      Matches:      37
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      82.08%      Indels:      0
DB:              9      Gaps:      0
```

US-09-997-610-2\_COPY\_1\_45 (1-45) x HS302D9 (1-145880)

```
QY      1 11eValVal11leProVal11leProVal11leThra1aVal11leg1uH1sVal1gluVal1a1a1g1Pro 20
Db      34786 ATAGTGCATACCTGCTTATATACGCGACATGATGACATGTGAAGTTGCTGACACT 34845
QY      21 Pro1aH1sPro1a1g1Pro1g1u1a1g1Pro1g1a1a1Pro1g1y 37
Db      34846 CCAGCAGACCCCGAGGCCCCGAGAGAGAGTGGGCGCTCTGTCAGCAGGT 34896
```

```
RESULT 2
AF017777
LOCUS
DEFINITION
AF017777 66669 bp DNA linear INV 04-AUG-1998
Drosophila melanogaster tweety (tly), flightless (fli), dodo (dod),
penguin (pen), small optic lobes (sol), innocent bystander (iby),
wacław (waw), bobby sox (bbs), slugish (slg), helicase (hlc),
misato (mat), and la costa (lcs) genes, complete cds.
AF017777 U00043 U28044
AF017777.1 GI:3004652
Drosophila melanogaster.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephyridioidea; Drosophilidae; Drosophila.
1 (bases 1 to 66669)
de Couet,R.G., Fong,K.S., Weeds,A.G., McLaughlin,P.J. and
Miklos,G.L.
Molecular and mutational analysis of a gelsolin-family member
encoded by the flightless I gene of Drosophila melanogaster
Genetics 141 (3), 1049-1059 (1995)
96129280
PUBMED
JOURNAL
MEDLINE
8582612
2 (bases 1 to 66669)
Miklos,G.L., Yamamoto,M., Burns,R.G. and Maleszka,R.
An essential cell division gene of Drosophila, absent from
Saccharomyces, encodes an unusual protein with tubulin-like and
myosin-like peptide motifs
Proc. Natl. Acad. Sci. U.S.A. 94 (10), 5189-5194 (1997)
9144213
PUBMED
JOURNAL
MEDLINE
97289742
3 (bases 1 to 66669)
Maleszka,R., de Couet,R.G. and Miklos,G.L.
Data transferability from model organisms to human beings: insights
from the functional genomics of the flightless region of Drosophila
Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3731-3736 (1998)
98188272
PUBMED
JOURNAL
MEDLINE
9520435
4 (bases 1 to 66669)
Miklos,G.L.G., Kasprzak,A., Mason,J., de Couet,H.G., Hayward,D.,
Hall,D., and Maleszka,R.
Direct Submission
Submitted (04-AUG-1997) Visual Sciences, Research School of
Biological Sciences, The Australian National University, Sullivan
Creek Rd., Canberra, ACT 0200, Australia
On or before Apr 2, 1998 this sequence version replaced gi:2098624,
gi:1052878.
FEATURES
source
gene
mRNA
CDS
Location/Qualifiers
1..66669
/organism="Drosophila melanogaster"
/strain="CantonS"
/db_xref="taxon:7227"
/chromosome="X"
complement(1639..7694)
/gene="tly"
complement(join(1639..2707,2789..2800,3622..3791,
3917..4680,4745..5230,5525..5809,7521..7694))
/gene="tly"
/product="tweety"
/codon_start=1
/product="tweety"
/db_xref="GI:3004653"
/translation="MGDYHEFTDQYKVPYAKILALPHYNITFRKINSFRPDELY
LESIGIGSVPAALIIYSLGLFLYIMTRCCDRKPRRAHSTLSKVALSYTVWCAG
IGGLYGNDDILHNLLEVLTRGKRVNDLVTTIRQTHILNTLTNRPLQVLELADIF
DQPVSNQTAISKPLFSLNIVOGNVTATATNAASDIRPLMGISMTNHLTRGDQWELFM
PGVTATLALILVLCALVIVGARRSCALILFVSCGLAATVGSWLMGSLYSSVAAG
DLTSPADFLVSTAPRDLPTVLLHTYQCERGHNPTPTORRESQNSLNARSMATV
```

MKLSILVFFSSQSDLPKRLCAVNAADNNSERLLDYLQVCKRHVHNPLIAPARGLCEGQ  
 LGVLVTLIASPIALITLLTVAWDSHTSYIEKRDVYQDVEPSITSLPAARQHCME  
 MNAARLPENRHHGHPSPVIGSHTLDPHSKROCHENMAHAITQOMKAMSTHTLGSL  
 PSYNSHPHPTHPHPHQAHAANAANMPPTTQAQOQOQOQOQOQOQOQOQOQOQOQOQ  
 QPITCHHPHPHPHPHQHSHSHSAVAVAAYOHQAIYHQOQAOQOQGTYYTAHAHAHA  
 HHLGQSOQYIOQOYIQAHLAPOLANGNPSHTYPLVAVSQGSITYNSNATKRNQSGH  
 GPQIPAHQHPSPHQOQOQOQPPSPQOQOQOQHOLKSPQHQOQOQOQOQOQOQOQOQOQ  
 HHQOQ  
 KPSILMCPLELPKQADPKVESITYEORQOAHQNTSK\*  
 complement(5809. .7521)  
 /gene="tly"  
 8586. .13783  
 /gene="fli"  
 join(8586. .8705,8869. .12023,12260. .12410,12533. .13783)  
 /gene="fli"  
 /product="flightless"  
 join(8652. .8705,8869. .12023,12260. .12410,12533. .12943)  
 /gene="fli"  
 /codon\_start=1  
 /product="flightless"  
 /protein\_id="AAC28407.1"  
 /db\_xref="GI:3004660"  
 /translation="MSVLPFVARGDFTKNDSPSPSSRQMSRWOMLTLDPTQLEAE  
 PELGLQQLQHLSSLNHRLEKIGELTELTELSCRLSDLNHQNKSGIIPPLPLAE  
 TTYLDDSHNKLKVPGELEFRANKLIVLMSNNQLESTPYPLFHLTDLFDLSHNRLE  
 LTPDQTRLINKTLTDLSHNLEPLQRLPSQSLVFLKMSGTQRTLLNPTSIDSL  
 ANCELDLSHNSLPKLPDCCVNVYTLVRLMSDNELTETAGVELMORLESINLSRQ  
 LVLPALALCKLRLRLLVNDNKNLNPFGISGIGALVEVSAANNLLEWPGILCR  
 GVALQOLNCSKRLTLTLDPAIHLEGGDQDLENNELWMPKPESAKATSEFVNT  
 DPLQIOLQLACAAVPSMPSATPKDXTARKRLRGRSGSGQDLATVILKGMQVAA  
 KQKDNAGAVPEDGKPESLKRRNDESLERQJDIKFEFKDQDQGLTITWELIENLT  
 PKNIEEVGKRYEGDCYTVLTKFIDLGLDWEIFPMWGNATLIDKRAACAIAHVL  
 RNLGALGATVAREEDDEBPQLFETFEVITYEGRTATGYTTEEMIHITRLYLVH  
 AYATHTLPRAVATTSDDPSHAFVLDLGHITYIMGSESKTILMSKRLAEKISTMTY  
 ERNKCEIOLEQGESEAEFMQGLGMYTSESDAABPKCHEVEDQPOPLRYOVQGL  
 MGYLELPQYELQKLCHTLILNSKHVYITLCYTDLPVWGKSGSTLYRAAAYKLSREL  
 FNMADPDTALVWRYPESGNEQYITRTFAQMDVMAVDTRAKSVAKTGANLTQMAR  
 QOETRIDLALPMPROSAMPLAEAEQLEEBMNVDLEMMAEVLENNKKEVRLPESELGR  
 FYEGEYVILCYRCYPIEIEBPNGSEGDANPAADVSSANNOPEDEIQCVYFMQGRRA  
 AGMKWGLPTEFTLQKKFAMGEELEVARYFEQOENILKMSHKKRTIHTGKKRDAKA  
 HTXKGSPEVFEPLHNSGALITTRLIOLINPDVAHLSACYIILHVPFEETEDSOSGIV  
 YWVIGSKACNEAKIYODIAEOMENSPWVSLQDILNGDEAPENFWVALGSKRPDITRA  
 EYVNTIRLRCSENGEYITYVAKEDACDQDDLITDINGEYFIIMQGRCSSEVDA  
 VKIAYVSAOYVYQHRIRIKQPERPKLFLTYKNKESRFTKCFHMSAFKRYL"  
 14131. .15603  
 /gene="dod"  
 join(14131. .14252,14364. .14696,15066. .15603)  
 /gene="dod"  
 /product="dodo"  
 join(14195. .14252,14364. .14696,15066. .15175)  
 /gene="dod"  
 /codon\_start=1  
 /product="dodo"  
 /protein\_id="AAC28408.1"  
 /db\_xref="GI:3004661"  
 /translation="MPDAEQLPDGMEKRTSRSTGMSYILNMYTKESQMDPTAPAKRT  
 GGSAGGACAPDEVCILHLVYKHGSRSPSWMEANITTKBEADLLLEVYRNKIVQOQ  
 EATFDELAISYSCSSAKRGGDLGKFGCGMQQAEDAAFKLVNQDLSGIVDSHGLH  
 IILRKA"  
 complement(15247. .18587)  
 /gene="pen"  
 complement(join(15247. .15493,15707. .16182,16281. .16683,  
 16752. .17454,18055. .18587))  
 /gene="pen"  
 /product="penguin"  
 complement(join(15308. .15493,15707. .16182,16281. .16683,  
 16752. .17454,18055. .18500))  
 /gene="pen"  
 /note="contains pumilio repeats"  
 /codon\_start=1  
 /product="penguin"  
 /protein\_id="AAC28403.1"  
 /db\_xref="GI:3004656"

```

/translation="MSSSEKRGAPGANAASPLKLHSSGSKANQKPKFGAGAGAK
PKKPKGKSGSEVKGAGGAKKDKKSGPGGPKFKSGKGLGKGGKPKGPGKPOPD
QAPRAGEKQDNMKFKKKKKDLKKRSADQTEIKEANOJHEKRCRPTNKKRYVE
QIYKLANGLDTISKVKAHDJARVIOSMILKTAISPAKLBESKYLTPFVEKQSYAKO
FCVQMKLKYGPATKRAKLVDSLVGHIVYRAGHSISGLSDYSKYOATPNQRTVYRGE
YDYLDRKAKDSNVKTLSDTYKFEATNNKASTLSGVANDHVNKOYDVALIKVHKEH
YLRACDEDEEKEETVAFALVPIMLSRKESSEAAVJCEFYKSTPKNSRAIKVIMKE
LLKIANHGHVPLISLNLMDTKATKATKIDYHLDGOKALMSSPYCRATYOMLVAP
GDTTCFPHFEITRYVEEGLAKGKKEKRLKEILDEQLEAPVIAQSIADPAKFNQVSLI
LVYGOILNHIQGESYEKAASALAQVNVQPEMRISADAPQPODKKRNQVNSVIAQ
APKQKREKLLVYESSDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE
PAKKEEPAALPVGSEIEGEMHIVLKIKLNKGKEGEPFSOOLLQNLNLSVYLKAWIG
VNRACFVLKXIVECPALDDCKKATIAEERSLSQLADBKPTGAKLAAKIDICK"
19098..26653
/gene="sol"
/locus="19098..19602,20609..21234,21502..22414,22843..231
21238..24826,24899..25405,25465..25731,25803..25945,
26004..26653"
/gene="sol"
/product="small optic lobes"
/mol_weight=19361..19602,20609..21234,21502..22414,22843..231
21238..24826,24899..25405,25465..25731,25803..25945,
26004..26181"
/gene="sol"
/codon_start=1
/product="small optic lobes"
/protein_id="AAC28409.1"
/db_xref="GI:3004662"
/translation="MGTISSVLOWSCNKTINPTESLKCFCNGCTVRKVPFQOQOQO
RSSSTJATWADDALEBOEAKEOBERKEKGAARVAASRYKYVUSYLRGLCKPQKN
SONLPANCVDCEPTDRYKIKYSIELYRHFNSPALNRMYCNACGDNSSVTHGCLIDPT
VSTLAPYTKDAIADRGQDLASGLNKGDELADNHSRHHNNHNYIMOLEEQRDQHLH
SOHLKRLHKGSAAGSGSGSGSGSGSLRRTQSLSTADKASGRCHCVANNQSKDIT
ENLPILKIPAPDLGTIPVPAACSNMRFALNDYCFRQKQNNNNKNNHNVYVESGAKRK
YNEFTTILSRSAADAGHGOQKPLQVNVNLNLLOEPOKSPAPROLOQKTPREBA
ATSNMPQFTIPRNGVFIAYNEMSEPMASSSVSNHHNNHNSNSNSGNSNITL
NNSSSSSGSKLLENCEVALAOQQLRAAADAQAATVALASPEAKMAPRAT
ATMPYIAOVNKHKLKTKQOIASBSOTNNNTGELADVDSYAGTASGISTDSGGA
SSESQVEHESIYAKVWKPGRKATSKIMHPDSSSRISGAASAAGTASGALAAAVN
GAAASRHDNKTOLGNGSRKMWICISGAANRMLDTCEMCEKAAEQOQOQOQLOQO
OQOQOQOHHNNHHNHHLOOQOEAARDPEMTCCKTLYNSTMACVVGSGSLKSISS
JEDMLRKKEGFEKTSHTCINLSPVSCSKNHROPQSMEMEVRRPQSGEEO
DMAAVGGGSGSAHOSGANEKYKAPRLNPLTSTALPMPLOLPTSTAGLGGSSPSP
RQQLPLLOQDRNSSGQALPKRHSFGSLTPENKSTIAGLAVNLQOGQCGSASVYS
AAGGSGAGAVGASSTSKWQCPACTYUQNCASVVCIDCSSPGRGLASVADLGLGRS
VVALTPADIDROESKLEMNROLTEETALIKMWNIIQYCRDSELTVDSPFPAPKSL
YNNPSSGEGNPPVOWMRPHREINIDCGAVAPPAVRPEPLDIOGVLGCMCLSLSL
AYLARRELDKEVLTVEIKGOGAYOYRLCKQKWTVLVDDLPCKDRGLYVISOAK
RKQLVPLTEKAVAKIHCEYALVSGALIEGLATILGLACBESIPQASLPMPEDELT
DKLILMAOQLSSRCVFLMGASGCGGNMKADEEETQOKLYRHNATSVLDKVDLQGHR
LKLNLPMNGHYSMRGWSDSDSLMTDLDRLALPHGASGVWISFEVLYKDFPICDI
CYVRGASNEVEFLQTLQPLCSISCVLTLYTEPEAFPLFQEGNSKSEKSORLODLC
VIFPFRSAPAREIGRLVESHKROVGFVGCHEKLMERDIYLLVCLAFMHNHGTIEDPH
OYPCOIIAHSKRLIVEQISPRPILAAATISLTGQORHREGMCAVYLKGNAAV
GLVAWENRHNKNTIHKYCDQOESYNNVSTRQELKVDSPVPLQOVIITVLTQLES
GSLIAHLRHLRANSRGLHDWGPAGATPCPIENVHGLAPRLT"
27035..30414

```

REFERENCE  
AUTHORS

1 (bases 1 to 171548)  
Celniker,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,  
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,  
Rogers,Y., An,H., Baldwin,D., Bonzon,J., Beeson,K.Y., Busan,D.A.,  
Carlson,J.W., Center,A., Chame,M., Davenport,L.B., Dietz,S.M.,  
Dodson,K., Dorsett,V., Doull,E., Doyle,C., Dresnek,D., Farfan,D.,  
Fierlier,S., Frisze,E., Gallie,R.F., Gary,N.S., George,R.A.,  
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,  
Idegawa,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,  
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,  
Pacled,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,  
Phoumenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,  
Stapleton,M., Strong,R., Svitskas,R., Tector,C., Williams,S.M.,  
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.

TITLE  
JOURNAL  
AUTHORS

Sequencing of Drosophila chromosome X, region 20B-20C  
Unpublished  
2 (bases 1 to 171548)  
Celniker,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,  
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,  
Rogers,Y., An,H., Baldwin,D., Bonzon,J., Beeson,K.Y., Busan,D.A.,  
Carlson,J.W., Center,A., Chame,M., Davenport,L.B., Dietz,S.M.,  
Dodson,K., Dorsett,V., Doull,E., Doyle,C., Dresnek,D., Farfan,D.,  
Fierlier,S., Frisze,E., Gallie,R.F., Gary,N.S., George,R.A.,  
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,  
Idegawa,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,  
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,  
Pacled,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,  
Phoumenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,  
Stapleton,M., Strong,R., Svitskas,R., Tector,C., Williams,S.M.,  
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.

TITLE  
JOURNAL

Submitted (12-JUL-2001) Berkeley Drosophila Genome Project,  
MS 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,  
Berkeley, CA 94720, US

COMMENT

Sequence submitted by:  
Lawrence Drosophila Genome Project  
Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720

FEATURES  
source

This sequence was assembled using end sequences from a whole genome  
shotgun and from subclones of this BAC and its neighboring clones.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdp@fruitfly.berkeley.edu](mailto:bdp@fruitfly.berkeley.edu).  
Location/Qualifiers  
1..171548  
/organism="Drosophila melanogaster"  
/strain="Y; cn bw sp"  
/db\_xref="taxon:7227"  
/chromosome="X"  
/map="20B-20C"  
/clone="BACR23J18 (DI450)"  
/clone\_1lb="RPCI-98 (Roswell Park Cancer Institute  
Drosophila melanogaster BAC library, partial EcoRI in  
pBACE3.6)"

BASE COUNT 47824 a 36178 c 37366 g 50180 t

ORIGIN

Alignment Scores:

Pred. No.: 1.11e+03 Length: 171548

Score: 89.00 Matches: 21

Percent Similarity: 47.17% Conservative: 4

Best Local Similarity: 39.62% Mismatches: 12

Query Match: 37.08% Indels: 16

DB: 3 Gaps: 2

US-09-997-610-2\_COPY\_1.45 (1-45) x AC0092494 (1-171548)

OY 2 ValVallelpeVallleuileThralaValalleluHuhsValgluaValala----- 18  
|||||::: |||||::: || :: |||||::: ||  
Db 51626 GTCTGTGGAGAAAGCTTGTTCAGACGCCTCAATTTTC---GTGAGATGTGCTTTGGTTCGAT 51682

Y 19 -----GlyProProAlaniSProAlyPro 26

Db 51683 GGAGGTCCCATGCTTGGGCGCTCTGTCACACCTGCTCTCTGCGACATCTGCTCT 51742  
Qy 27 ProgluGlValG1yProProglYAlaProglYLeuPro 39  
||||| ||||||| ||||| |||  
Db 51743 CCTGGGCGCTCCTGCTGCTCTGCTCTGCTCTGCGCTCTCT 51781  
RESULT 5  
AC011251  
LOCUS  
DEFINITION  
AC011251 173508 bp DNA linear INV 07-JUN-2001  
Drosophila melanogaster, chromosome X, region 19F-20A, BAC clone  
BACR09F10, complete sequence.  
AC011251  
AC011251.5 GI:14327739  
HTG.  
SOURCE  
ORGANISM  
Drosophila melanogaster.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
AUTHORS  
Celniker,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,  
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Boush,D.A.,  
Rogers,Y., An,H., Baldwin,D., Bonzon,J., Beeson,K.Y., Busam,D.A.,  
Catalson,D.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M.,  
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dreese,D., Fartan,D.,  
Fertler,S., Frisoe,E., Galle,R.F., Garg,N.S., George,R.A.,  
Gonzalez,M., Houch,J., Hoskins,R.A., Hostin,D., Howland,T.J.,  
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,  
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nuno,J.,  
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,  
Phouanavong,S., Piltman,G.S., Puri,V., Richards,S., Scheeler,F.,  
Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M.,  
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.  
Sequencing of Drosophila chromosome X, region 19F-20A  
Unpublished  
2 (bases 1 to 173508)  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,  
Butenheff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
Doyle,C.M., Fartan,D.E., Galle,R., George,R.A., Harris,N.L.,  
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,  
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,  
Pfeiffer,B., Poon,U., Sequeira,A., Sehl,H., Shtir,E.,  
Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
Rubin,G.M.  
Direct Submission  
Submitted (05-OCT-1999) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA  
On Jun 7, 2001 this sequence version replaced gi:6087907.  
COMMENT  
Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720  
This sequence was assembled using end sequences from a whole genome  
shotgun and from subclones of this BAC and its neighboring clones.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu).  
FEATURES  
Source  
Location/Qualifiers  
1..173508  
/organism="Drosophila melanogaster"  
/strain="Y; cn bw sp"  
/db\_xref="taxon:7227"  
/chromosome="X"  
/map="19F-20A"  
/clone="BACR09F10 (D1128)"  
/clone\_lib="RP1-98 (Roswell Park Cancer Institute  
Drosophila melanogaster BAC library, partial EcoRI in  
PBAC3.6)"  
BASE COUNT 47977 a 37003 c 38363 g 50165 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.12e+03 Length: 173508  
Score: 89.00 Matches: 21  
Percent Similarity: 47.17% Conservative: 21  
Best Local Similarity: 39.62% Mismatches: 12  
Query Match: 37.08% Indels: 16  
DB: 3 Gaps: 2  
US-09-997-610-2\_copy\_1\_45 (1-45) x AC011251 (1-173508)  
Qy 2 ValVal11eProVal1eul1eHrAlaVal11eG1uH1SVAlG1uValAlA----- 18  
|||||: |||||: ||: |||||  
Db 91739 GTCGTGTGGAAGTCTGTGTGAGCCCTCAGTTGTC--GGGATGAGTGTGGTTCGAT 91795  
Qy 19 -----G1yProProAlaH1aSPARgPro 26  
||||| ||||| |||  
Db 91796 GGAGTCCCATGCTTGGGCGCTCTGTCACACTGCTCTCTGCGACATCTGCTCT 91855  
Qy 27 ProgluGlValG1yProProglYAlaProglYLeuPro 39  
||||| ||||||| ||||| |||  
Db 91856 CTTGGGCGCTCCTGCTCTGCTCTGCTCTGCGCTCTCT 91894  
RESULT 6  
AE003568  
LOCUS  
DEFINITION  
AE003568 292911 bp DNA linear INV 05-OCT-2000  
Drosophila melanogaster genomic scaffold 14200001386037 section 5  
of 5, complete sequence.  
ACCESSION  
AE003568 AE002620  
KEYWORDS  
SOURCE  
HTG.  
Drosophila melanogaster.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
AUTHORS  
1 (bases 1 to 292911)  
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,  
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,  
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,  
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,  
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,  
Man,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor  
Miklos,G.L., Abill,J.F., Agbayani,A., An,H.J.,  
Andrews-Pfankuch,C., Baldwin,D., Ballou,R.M., Basu,A.,  
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,  
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,  
Botchan,M.R., Bouck,J., Brokstein,P., Brothier,P., Burtis,K.C.,  
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,  
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de  
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,  
Dodson,K., Doup,L.E., Downes,M., Duncan-Rocha,S., Dunkov,S.,  
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Fertler,S.,  
Fleischmann,W., Foster,C., Gabrielian,A.E., Garg,N.S.,  
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gottrell,J.H., Gu,Z.,  
Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,  
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,  
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,  
Kensimon,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,  
Krivtsov,S., Kulp,D., Lai,Z., Lascko,P., Lei,Y., Levitsky,A.A.,  
Li,U., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,  
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Modyarty,C.,  
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,  
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,  
Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Piltman,G.S., Pan,S.,  
Pollard,J., Puri,V., Reese,M.C., Reinert,K., Remington,K.,  
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,  
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,  
Stapleton,M., Strong,R., Sun,E., Svirskaas,R., Tector,C., Turner,R.,  
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,  
Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,  
Mortley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,  
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,



University, Genome Sequencing Center, 4444 F  
Louis MO 63110 USA

```
/product="hypocretin1 protein 151h/c.8"
/protein_id="AAK39332.1"
```

```
/product="hypocretin protein 131h/c.8"
/protein_id="AAK39332.1"
```

[illegible]

<http://www.wormbase.org/db/seq/sequence?name=Y51HC.9;class=sequence>  
 complement(join(7941..7992,8063..8169,8301..8422,8469..8610,8657..8733,9106..9211,9269..9619,9674..981985..9954,9999..10117,10177..10213))  
 /gene="Y51HC.9"  
 /note="coded for by the following C. elegans cDNAs:  
 YK362g7.3, YK362g7.5"  
 /codon\_start=1  
 /product="Hypothetical protein Y51HC.9"  
 /protein\_id="AAK39340.1"  
 /db\_xref="GI:13775532"

**Alignment Scores:**  
 Pred. No.: 692 Length: 68661  
 Score: 87.50 Matches: 17  
 Percent Similarity: 58.82% Conservative: 3  
 Best Local Similarity: 50.00% Mismatches: 9  
 Query Match: 36.46% Indels: 5  
 DB: 3 Gaps: 1

US-DB-997-610-2\_COPY\_1\_45 (1-45) x AC024805 (1-68661)

OY	17	ValAlaGlyProProAlaHisProArgProGluGluValAlaGlyProProGlyAlaPro	36
Db	66661	GTACCTGACCTCCAGGGGTTCTGACCTCCGGGTACTGACCTCGGGGTACT	66602
OY	37	GlyLeuProGlnTyr-----ThrGlyGluLeu	45
Db	66601	GGACCAACCCGAAATACCGTAAGAACCTGCACTGGCCAAATTA	66560

**RESULT 8**  
 AC006799/c 278007 bp DNA linear HTG 23-FEB-1999  
 LOCUS  
 DEFINITION  
 7 unordered pieces.  
 AC006799  
 AC006799.1 GI:4263139  
 HTG: HTGS\_PHASE1.  
 CAenorhabditis elegans.  
 CAenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.  
 1 (bases 1 to 278007)  
 Waterston,R.H.  
 The sequence of *Caenorhabditis elegans* clone  
 Unpublished  
 2 (bases 1 to 278007)  
 Waterston,R.H.  
 Direct Submission  
 Submitted (23-FEB-1999) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA

**REFERENCE**  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

**COMMENT**  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 7 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

*	1	2255: contig of 2255 bp in length
*	2256	2259: gap of unknown length
*	2270	5536: contig of 3267 bp in length
*	5537	5550: gap of unknown length
*	5551	23191: contig of 17641 bp in length
*	23192	23205: gap of unknown length
*	23206	60448: contig of 37243 bp in length
*	60449	60462: gap of unknown length
*	60463	112471: contig of 52009 bp in length
*	112472	112485: gap of unknown length
*	112486	188654: contig of 76169 bp in length
*	188655	188668: gap of unknown length

FEATURES	* 188669	278007: contig of 89339 bp in length.
source	Location/Qualifiers	
	1..278007	
	/organism="Caenorhabditis elegans"	
	/db_xref="taxon:6239"	
	/clone="Y51H7"	
BASE COUNT	87340 a 51392 c 48889 g 90302 t	84 others
ORIGIN		
Alignment Scores:		
Pred. No.:	2.27e+03	Length: 278007
Score:	87.50	Matches: 17
Percent Similarity:	58.82%	Conservative: 3
Best Local Similarity:	50.00%	Mismatches: 9
Query Match:	36.46%	Indels: 5
DB:	2	Gaps: 1
US-09-997-610-2_COPY_1_45 (1-45) x AC006799 (1-278007)		
Qy	17 ValaAlaGlyProProAlaHisProAlaArgProProGluGluValAlaGlyProProGlyAlaPro 36	
Db	66563 GTACCTGGACCTCCAGGGGTTCTGTGACCTCCGGGGTACTGGACCTCTGTGGTACT 66524	
Qy	37 GlyLeuProGlnTyr-----ThrglyGluIle 45	
Db	66523 GGACCAACCGAATACCGTAGAACCGTCACTGGCCATTA 66482	
RESULT 9		
AX146422		
LOCUS	AX146422	4428 bp DNA linear PAT 31-MAY-2001
DEFINITION	Sequence 3 from Patent WO0134647.	
ACCESSION	AX146422	
VERSION	AX146422.1	GI:14284845
KEYWORDS		
SOURCE		
ORGANISM	cow.	
	Bos taurus	
	Eumetazoa; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;	
	Bovidae; Bovinae; Bos.	
REFERENCE	1 (bases 1 to 4428)	
AUTHORS	Bell, M.P., Neff, T.B., Polarek, J.W. and Seeley, T.W.	
TITLE	Animal collagens and gelatins	
JOURNAL	Patent: WO 0134647-A 3 17-MAY-2001;	
FEATURES	FIBROGEN, INC. (US)	
source	Location/Qualifiers	
	1..4428	
	/organism="Bos taurus"	
	/db_xref="taxon:9913"	
BASE COUNT	973 a 1185 c 1400 g 870 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	91.6	Length: 4428
Score:	86.00	Matches: 16
Percent Similarity:	61.54%	Conservative: 0
Best Local Similarity:	61.54%	Mismatches: 10
Query Match:	35.83%	Indels: 0
DB:	6	Gaps: 0
US-09-997-610-2_COPY_1_45 (1-45) x AX146422 (1-4428)		
Qy	18 AlaAlaGlyProProAlaHisProAlaArgProProGluGluValAlaGlyProProGlyAlaProGly 37	
Db	538 GCTGCTCTCTCTGGCCACCCGACCCCTCTGGCAATCTGGCAATCTCTGCTGGCCCTGGC 597	
Qy	38 LeuProGlnTyrThrGly 43	
Db	598 GCTCCAGATACCAAGT 615	
RESULT 10		
AX146424		
LOCUS	AX146424	4428 bp DNA linear PAT 31-MAY-2001
DEFINITION	Sequence 5 from Patent WO0134647.	

```

ACCESSION      AX146424
VERSION        AX146424.1  GI:14284846
KEYWORDS
SOURCE
ORGANISM
  Bos taurus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
  Bovidae; Bovinae; Bos.
REFERENCE
  Bell,M.P., Neff,T.B., Polarek,J.W. and Seeley,T.W.
  Animal collagens and gelatins
  Patent: WO 0134647-A 5 17-MAY-2001;
  FIBROGEN, INC. (US)
FEATURES
  source
    1..4428
    /organism="Bos taurus"
    /db_xref="taxon:9913"
BASE COUNT    974 a 1185 c 1399 g 870 t
ORIGIN
Alignment Scores:
Pred. No.:    91.6      Length:    4428
Score:        86.00     Matches:    16
Percent Similarity: 61.548   Conservative: 0
Best Local Similarity: 61.548   Mismatches: 10
Query Match:  35.838           Indels:    0
DB:           6              Gaps:    0
US-09-997-610-2_COPY_1_45 (1-45) x AX146424 (1-4428)
QY      18  ALaGcYPrOPrOAlAhSPrOaGrPrOPrOGUgUvAlGyPrOPrOGUyAlAPrOgLy 37
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      538  GCTGTCCTCTCTGGCCGCCAGCCCTGTCGACACTGTGGCATTCTGTCGTCCTGGCC 597
QY      38  LeuPrOGInTyTrHrGly 43
          |||  |||  |||  |||
DB      598  GCTCCAGATATACCAAGCT 615
RESULT 11
AV052763      5358 bp      mRNA      linear      VRT 17-JUN-20022
LOCUS
DEFINITION    Xenopus laevis type XVIII collagen short variant mRNA, complete cds.
ACCESSION    AY052763
VERSION      AY052763.1  GI:21434788
KEYWORDS
SOURCE
  ORGANISM
    African clawed frog.
    Xenopus laevis
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae;
    Xenopodinae; Xenopus.
  1 (bases 1 to 5358)
  Elamaa,H., Peterson,U.T., Pihlajaniemi,T. and Destree,O.
  Cloning of three variants of type XVIII collagen and their
  expression patterns during Xenopus laevis development
  Mech. Dev. 114 (1-2), 109-113 (2002)
  21993083
  JOURNAL
  MEDLINE
  PUBMED
  11996984
  2 (bases 1 to 5358)
  Elamaa,H., Peterson,U.T., Pihlajaniemi,T. and Destree,O.
  Direct Submission
  Submitted (28-AUG-2001) Collagen Research unit, Bloccentr and
  Department of Medical Biochemistry, University of Oulu, P. O. Box
  5000, Oulu 90014, Finland
  Location/Qualifiers
    1..5358
    /organism="Xenopus laevis"
    /db_xref="taxon:8355"
    1..3924
    /codon_start=1
    /product="type XVIII collagen short variant"
    /protein_id="AAU14257.1"
    /db_xref="GI:21434789"
FEATURES
  source
    1..5358
    /organism="Xenopus laevis"
    /db_xref="taxon:8355"
    1..3924
    /codon_start=1
    /product="type XVIII collagen short variant"
    /protein_id="AAU14257.1"
    /db_xref="GI:21434789"

```

/LienstallpknmlrCmWmVtLVAASHVtLSSODKQDGEVGLLOJIGEPVW  
 EITVTPEDPtknrvpVDEKQKANTQVOMVHLPMDNRDLSLEHFOVSTNKNGKJLFIATIT  
 DASOVIYICVtLSEYKCDKODJLFIYTEREYKQVAGVAGADPDTQVILADLVANK  
 DEBVLVtLPMODELLEHFRFESDPEMLLEEGSLPEAOVAGADPDTQVILADLVANK  
 PRAEROCEDDEDELEHFRFESDPEMLLEEGSLPEAOVAGADPDTQVILADLVANK  
 EIVVtGQVQOQJIKXGEMQKKEKERLTPKDSVSAZSPASISSEKQKQKEVCSK  
 AGFVPPKQKQDQKPGMGRPRGAPGPGSILLOKSTGTSVRVtGPPQBPQAPACK  
 DGEFDEDEDEGKATGEVQPOCFETPEDPKCEKQDTCGMHNGPGLGPPGPGAPK  
 TDKTLFIIDMEKSGSGOLEHVKCPPLPDPBPQVPLGPGVGFNGNLSLGPSSG  
 LPLGKGEJIBPGRPLGPGKNGLPGASVAGTGEQDLEJLPGACAKNPEQOD  
 PGAGQPGSLAGLPGMPKQBPQBPPEPPVRSGLTGDMEASGLQFLREBQOT  
 PGLPAGKQDEQNPOLPELOQKQKQSDPEPEBQKQGLDGPGEGLRGKGERGEGTET  
 GHVtGvIPGPPGPGQVLYPSDSKQSVAGATGPEPRQGLIQLPBPQPKQDOJLIDF  
 GHTGvAGKGEKPELMLRNPQJLTYPDSXGEMQDQKRPASVQKQKEJLIDF  
 PGRGQVNGVtLXGEMQKQSSGSLGPPGPGPPGPGSVENDOMIPAYGQKNGE  
 PPGQGVQVtGAPGPPGQVLTSTFALVREKGLKXGESSGAGVQIGPPGNGG  
 FPLGPGKAGSGQVPEPVPPEPSGJTEEGHQRNHPGPPGLGASVtGPHROTQ  
 S1TGPPGKQPGPQTSQSGAULOTHOTMLNAPRLDEGTVLQVORALVYKQ  
 GFRJLILEYVtLSSRODNEVAAEAPVVPYPRKXSDJLOSTEENSNNANPSCA  
 RNPADPLPEPRPDPDPDHQNGHFRLEPSVtHODFRPALTVALPALSISGM  
 K1NGVQFQDEBARKGLGHTFRALSRLQDLSIVYRRADROSVLQVNLRLDVLV  
 NMDSLFGSQAQSNQSGARLISFPGKQVNDPMPKQKMWMSDAGGRtLTSYCEYWR  
 NDEBAEVtGQASLTSGLQLEORQSNKNNIYLCIENSEFNTNKR"

Alignment Scores:	
Pred. No.:	119
Score:	85.50
Percent Similarity:	62.50%
Best Local Similarity:	56.25%
Query Match:	35.62%
DB:	5
Gaps:	5
Length:	5358
Matches:	18
Conservative:	2
Mismatches:	7
Indels:	2
Gaps:	2

US-09-997-610-2\_COPY\_1\_45 (1-45) x AY052763 (1-5358)

QY 12 ILEGUHSVALGLVALAGLYPROPROALAHISPROARGPROPROGLUGLVALGLY 31

Db 1402 TTGGAGCATGTCAG-----GGACCCCTGGTTGGCCTGGGCCCA-----GGT 1446

32 PROPGGLYALPROGLYLEUPROGLINTYRTHRGly 43

DB 144 / CCCTCAGGTCGTACCTGGGTACCCGGAGTACCCCGA 1482

AC079422

LOCUS	AC013422	25841 bp	DNA	linear	HIG 01-SEP-2000
DEFINITION	Mus musculus chromosome 16 clone RP23-109A22, WORKING DRAFT				

ACCESSION	SEQUENCE, or unordered pieces
AC079422	

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

ORGANISM

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS DOE Joint Genome Institute

**JOURNAL** Unpublished

**AUTHORS** DOE Joint Genome Institute  
**TITLE** Direct Substitution

JOURNAL  
Submitted (01-SEP-2000) Production Sequencing Facility, DOE Joint  
Genome Institute 3800 Mitchell Drive Walnut Creek CA 94508 USA

COMMENT: -----Genome Center  
Center: Joint Genome Institute

Center code: JGI  
Web site: <http://www.tai.doe.gov>

Project Information

Center clone name: RPCI-23\_109A22

Summary Statistics

Consensus quality: 161724 bases at least Q40

Consensus quality: 191966 bases at least Q30

Consensus quality: 205448 bases at least Q20

Estimated insert size: 220300; agarose-gel estimation

Estimated insert size: 250317; sum-of-contigs estimation

Quality coverage: 3.11 In Q20 bases; agarose-gel estimation

Quality coverage: 2.74 In Q20 bases; sum-of-contigs estimation

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 62 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

*	1	101:	contlg of 1101	bp in length
*	1101	1201:	gap of unknown length	
*	1202	2412:	contlg of 1211	bp in length
*	2413	2512:	gap of unknown length	
*	2513	3675:	contlg of 1163	bp in length
*	3676	3775:	gap of unknown length	
*	3776	4879:	contlg of 1104	bp in length
*	4880	4979:	gap of unknown length	
*	4980	6106:	contlg of 1127	bp in length
*	6107	6206:	gap of unknown length	
*	6207	7322:	contlg of 1116	bp in length
*	7323	7422:	gap of unknown length	
*	7423	8536:	contlg of 1134	bp in length
*	8537	8656:	gap of unknown length	
*	8657	9938:	contlg of 1182	bp in length
*	9939	9938:	gap of unknown length	
*	11059	11058:	contlg of 1120	bp in length
*	11159	11158:	gap of unknown length	
*	11236	12325:	contlg of 1167	bp in length
*	12426	13425:	gap of unknown length	
*	13647	13746:	contlg of 1221	bp in length
*	13747	14871:	gap of unknown length	
*	14872	14971:	gap of unknown length	
*	14972	16434:	contlg of 1463	bp in length
*	16435	16534:	gap of unknown length	
*	16535	17812:	contlg of 1278	bp in length
*	17813	17912:	gap of unknown length	
*	17913	19230:	contlg of 1318	bp in length
*	19231	19330:	gap of unknown length	
*	19331	20400:	contlg of 1070	bp in length
*	20401	20500:	gap of unknown length	
*	20501	22256:	contlg of 1756	bp in length
*	22257	22357:	gap of unknown length	
*	22357	22977:	contlg of 1621	bp in length
*	22978	24077:	gap of unknown length	
*	24078	25281:	contlg of 1204	bp in length
*	25282	25381:	gap of unknown length	
*	25382	27415:	contlg of 2034	bp in length
*	27416	27515:	gap of unknown length	
*	27516	28277:	contlg of 1762	bp in length
*	28278	29278:	gap of unknown length	
*	29278	33364:	contlg of 2987	bp in length
*	33265	33464:	gap of unknown length	
*	33465	33843:	contlg of 1379	bp in length
*	33844	33943:	gap of unknown length	
*	33944	36718:	contlg of 2775	bp in length
*	36719	36818:	gap of unknown length	
*	36819	38355:	contlg of 1537	bp in length
*	38356	38455:	gap of unknown length	
*	38456	39544:	contlg of 1089	bp in length
*	39545	39644:	gap of unknown length	
*	39645	42096:	contlg of 2452	bp in length
*	42097	42196:	gap of unknown length	
*	42197	44338:	contlg of 2242	bp in length
*	44339	44538:	gap of unknown length	

44539	46550	contig of 2112 bp in length
44651	46750	gap of unknown length
44671	46936	contig of 2186 bp in length
44937	49036	gap of unknown length
52141	52140	contig of 3104 bp in length
52241	52240	gap of unknown length
52447	54747	contig of 2507 bp in length
54748	54847	gap of unknown length
54848	55017	contig of 4170 bp in length
55018	55117	gap of unknown length
55118	62801	contig of 3684 bp in length
62802	62901	gap of unknown length
65902	66517	contig of 3616 bp in length
66518	66517	gap of unknown length
70275	70374	contig of 3657 bp in length
70375	70375	gap of unknown length
75540	75539	contig of 5165 bp in length
75640	75639	gap of unknown length
79310	79309	contig of 3670 bp in length
79410	79409	gap of unknown length
84690	84689	contig of 5280 bp in length
84790	84788	gap of unknown length
88911	88810	contig of 5021 bp in length
89911	89910	gap of unknown length
89911	95503	contig of 5593 bp in length
95504	95503	gap of unknown length
95604	100928	contig of 5325 bp in length
100929	101028	gap of unknown length
101029	105513	contig of 4485 bp in length
105514	105613	gap of unknown length
105614	110389	contig of 4776 bp in length
110390	110489	gap of unknown length
110490	114852	contig of 4363 bp in length
114853	114952	gap of unknown length
114953	120726	contig of 5774 bp in length
120727	120826	gap of unknown length
120827	127059	contig of 6233 bp in length
127060	127159	gap of unknown length
127160	133607	contig of 6448 bp in length
133708	133707	gap of unknown length
133708	139628	contig of 5921 bp in length
139629	139728	gap of unknown length
139729	147785	contig of 8057 bp in length
147786	147885	gap of unknown length
147886	154861	contig of 6976 bp in length
154862	154961	gap of unknown length
154962	160982	contig of 6021 bp in length
160983	161082	gap of unknown length
161083	167776	contig of 6694 bp in length
167777	167876	gap of unknown length
167877	175123	contig of 7247 bp in length
175124	175223	gap of unknown length
175224	183464	contig of 8241 bp in length
183465	183564	gap of unknown length
183565	191380	contig of 7816 bp in length
191381	191480	gap of unknown length
191481	200932	contig of 9452 bp in length
200933	201032	gap of unknown length
201033	207587	contig of 6555 bp in length
207588	207687	gap of unknown length
207688	216967	contig of 9280 bp in length
216968	217067	gap of unknown length
217068	229646	contig of 12579 bp in length
229647	229747	gap of unknown length
229747	241492	contig of 11746 bp in length
241493	241592	gap of unknown length
241593	256417	contig of 14825 bp in length
241593	Location/Qualifiers	

[illegible]



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 19, 2003, 14:57:18 ; Search time 2076.08 Seconds  
(without alignments)  
2088.706 Million cell updates/sec

Title: US-09-997-610-2\_COPY\_1\_149  
Perfect score: 790  
Sequence: 1 IVVIVLITAVIEHVEVAGP.....KEEISKQOSIQEVTWLLKA 149

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Ygapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO/spool/US09997610/runat\_10022003\_160823\_23819/app.query.fasta.1.1635  
-DB=genembi -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=plc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09997610.qcgn.1.1365\_@runat\_10022003\_160823\_23819 -NCPU=6 -ICPU=3  
-NO\_XLPRY -NO\_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_inu:\*  
19: em\_lnu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rod:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_sy:\*  
39: em\_hcgo\_hum:\*  
40: em\_hcgo\_mus:\*  
41: em\_hcgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	623	78.9	145880	9 HS302D9	Z82198 Human DNA s
2	330	41.8	1265	10 TMSHP20A	D12974 Tamias asia
3	291.5	36.9	5121	10 AB067779	AB067779 Tamias si
4	220	27.8	2005	10 TMSHP25	D12975 Tamias asia
5	169.5	21.5	1918	10 RNO131848	AJ131848 Rattus no
6	167.5	21.2	1385	10 TMSHP27	D12976 Tamias asia
7	167.5	21.2	2609	5 CHKX	M13496 Chicken typ
8	167	21.1	559	10 AB06781353	AB067813 Callosciu
9	163.5	20.7	2235	6 AX332258	AX332258 Sequence
10	163.5	20.7	2235	9 HSCOL8A1	X57527 Human COL8A
11	163.5	20.7	2415	4 RABCOLVITI	J05042 Rabbit type
12	162.5	20.6	2506	9 BC013581	BC013581 Homo sapi
13	162.5	20.6	117000	9 AC069222	AC069222 Homo sapi
14	162.5	20.6	145880	9 HS302D9	Z82198 Human DNA s
15	162.5	20.6	156922	2 AC120598	AC120598 Homo sapi
16	162.5	20.6	166992	2 AC022883	AC022883 Homo sapi
17	162.5	20.6	182978	30 AC067824	AC067824 Homo sapi
18	161	20.4	7089	4 AF222861	AF222861 Sus scrof
19	159.5	20.2	1986	10 NMICOL8A	X53556 Bovine COL1
20	159	20.1	3143	4 BTCOL10A1	Z21610 M.musculus
21	158	20.0	3422	10 NMALTXCOL	X67348 M.musculus
22	158	20.0	9331	10 NMICOL10A	AC119229 Mus muscu
23	158	20.0	68834	2 AC119229	AC016087 Homo sapi
24	158	20.0	179303	2 AC016087	AC021709 Mus muscu
25	158	20.0	206213	2 AC021709	BC011061 Mus muscu
26	157.5	19.9	2435	10 BC011061	X63013 M.musculus
27	156	19.7	731	10 NMALTXCOL	X65120 H.sapiens C
28	156	19.7	1973	9 HS10A1COL	X72580 Homo sapien
29	156	19.7	3215	9 HSCOLX3	AX333243 Sequence
30	156	19.7	3226	6 AX333243	X60382 H.sapiens C
31	156	19.7	3326	9 HSCOLX1X	G28608 human STR S
32	156	19.7	3326	11 G28608	X98568 H.sapiens t
33	156	19.7	10058	9 HSCOLX	AL121963 Human DNA
34	156	19.7	107553	9 HS1336014	AL353573 Homo sapi
35	156	19.7	205594	2 AL353573	AF417206 Canis fam
36	155.5	19.7	582	4 AF417206	BC028770 Bos tauru
37	152.5	19.3	947	10 BC028770	AF269230 Bos tauru
38	152.5	19.3	1134	4 AF269230	AC125396 Mus muscu
39	152.5	19.3	200942	2 AC125396	AK074129 Homo sapi
40	151	19.1	4443	9 AK074129	AX330953 Sequence
41	151	19.1	4908	6 AX330953	AL138787 Human DNA
42	151	19.1	128133	2 AL138787	AC012141 Homo sapi
43	151	19.1	173817	2 AC012141	AX195207 Sequence
44	150.5	19.1	1152	6 AX195207	AX358517 Sequence
45	150.5	19.1	1152	6 AX358517	

RESULT 1

## ALIGNMENTS

HS302D9  
LOCUS HS302D9 145880 bp DNA linear PRI 12-DEC-1999  
DEFINITION Human DNA sequence from clone RPI-302D9 on chromosome 22 Contains  
GSSs, complete sequence.  
ACCESSION 282198  
VERSION 282198.2 GI:6572207  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS 1 (bases 1 to 145880)  
TITLE Bridgeman, A.  
JOURNAL Direct Submission  
Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
Requests: clonerequest@sanger.ac.uk  
On Dec 13, 1999 this sequence version replaced gi:3164067.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/projects/Celegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 22, constructed by the Sanger Centre Chromosome 22  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr22  
RPI-302D9 is from the library RPCI-1 constructed at the Roswell  
Park Cancer Institute by the group of Pieter de Jong. For further  
details see http://bacpac.med.buffalo.edu/  
VECTOR: pCYPAC2  
This sequence is the entire insert of clone RPI-302D9 The true left  
end of clone CTA-282F2 is at 69682 in this sequence. The true right  
end of clone CTA-415G2 is at 55167 in this sequence.

FEATURES  
source  
Location/Qualifiers  
1..145880  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="22"  
/clone="RPI-302D9"  
/clone\_1fb="RPCI-1"  
188..245  
/note="MER3 repeat: matches 144..209 of consensus"  
246..571  
/note="Alusx repeat: matches 1..312 of consensus"  
572..759  
/note="MER3 repeat: matches 1..144 of consensus"  
783..933  
/note="MER5A repeat: matches 26..187 of consensus"  
1033..1336  
/note="Alusp repeat: matches 1..299 of consensus"  
1450..1583  
/note="MIR repeat: matches 24..160 of consensus"  
1687..1752  
/note="L2 repeat: matches 2593..2661 of consensus"  
2350..2660  
/note="Alusc repeat: matches 3..309 of consensus"  
2684..2981  
/note="Alusq repeat: matches 2..300 of consensus"  
3323..3343  
/note="MIR repeat: matches 116..136 of consensus"

repeat\_region 3344..3652  
/note="Aluy repeat: matches 1..309 of consensus"  
repeat\_region 3653..3928  
/note="MIR repeat: matches 136..359 of consensus"  
repeat\_region 3929..4278  
/note="THE1B repeat: matches 3..364 of consensus"  
repeat\_region 4279..4485  
/note="MIR repeat: matches 359..568 of consensus"  
repeat\_region 5073..5176  
/note="52 copies 2 mer ct 78 conserved"  
repeat\_region 5181..5491  
/note="Alub repeat: matches 1..311 of consensus"  
repeat\_region 6369..6485  
/note="L2 repeat: matches 2579..2705 of consensus"  
repeat\_region 6647..6685  
/note="MAD1 repeat: matches 1..23 of consensus"  
repeat\_region 6686..6987  
/note="Alusx repeat: matches 1..302 of consensus"  
repeat\_region 6988..7036  
/note="MAD1 repeat: matches 23..77 of consensus"  
repeat\_region 7482..7754  
/note="Alub repeat: matches 9..290 of consensus"  
repeat\_region 7775..8060  
/note="Aluo repeat: matches 1..295 of consensus"  
repeat\_region 8414..8551  
/note="L2 repeat: matches 2553..2706 of consensus"  
repeat\_region 8914..9030  
/note="MIR repeat: matches 147..262 of consensus"  
repeat\_region 9110..9280  
/note="MIR repeat: matches 91..262 of consensus"  
repeat\_region 9283..9412  
/note="MIR repeat: matches 15..144 of consensus"  
repeat\_region 9521..9679  
/note="PAM repeat: matches 3..161 of consensus"  
repeat\_region 9820..10225  
/note="MSR repeat: matches 2..425 of consensus"  
repeat\_region 10179..10678  
/note="match: GSS: Em:B56592"  
repeat\_region 10204..10728  
/note="match: GSS: Em:AQ701486"  
repeat\_region 10249..10706  
/note="match: GSS: Em:AQ225495"  
repeat\_region 10312..10383  
/note="MIR repeat: matches 79..150 of consensus"  
repeat\_region 10718..11310  
/note="match: GSS: Em:B14024"  
repeat\_region 10784..11201  
/note="match: GSS: Em:B43656"  
repeat\_region 11838..11946  
/note="MIR repeat: matches 20..137 of consensus"  
repeat\_region 12174..12445  
/note="L2 repeat: matches 1988..2275 of consensus"  
repeat\_region 12444..12642  
/note="MIR repeat: matches 63..241 of consensus"  
repeat\_region 13017..13369  
/note="match: STS: Em:G49301"  
repeat\_region 13331..13397  
/note="MIR repeat: matches 174..244 of consensus"  
repeat\_region 13398..13698  
/note="Alusp repeat: matches 1..302 of consensus"  
repeat\_region 13699..13810  
/note="MIR repeat: matches 76..174 of consensus"  
repeat\_region 13806..13919  
/note="MIR repeat: matches 77..189 of consensus"  
repeat\_region 13945..14060  
/note="MIR repeat: matches 24..142 of consensus"  
repeat\_region 14061..14367  
/note="Aluy repeat: matches 1..301 of consensus"  
repeat\_region 14368..14452  
/note="MIR repeat: matches 141..225 of consensus"  
repeat\_region 14589..14679  
/note="MIR repeat: matches 173..262 of consensus"  
misc\_feature 14597..15201

```
/note="match: GSS: Em:AQ53482"
14616..15060
/note="match: GSS: Em:AQ370601"
14868..15040
repeat_region
/note="MIR repeat: matches 49..233 of consensus"
15071..15188
/note="L2 repeat: matches 2112..2239 of consensus"
15304..15399
repeat_region
/note="MIRB repeat: matches 1..99 of consensus"
15490..15662
/note="Alusg1 repeat: matches 2..114 of consensus"
15669..15727
repeat_region
/note="MIRB repeat: matches 119..178 of consensus"
15728..16027
/note="Alus repeat: matches 1..299 of consensus"
16028..16245
repeat_region
/note="MIRB repeat: matches 178..390 of consensus"
16346..16854
repeat_region
/note="AluY repeat: matches 1..300 of consensus"
18296..18323
repeat_region
/note="MSTR repeat: matches 2..29 of consensus"
18324..18392
repeat_region
/note="MER66-internal repeat: matches 4919..4993 of consensus"
18393..18712
repeat_region
/note="AluDb repeat: matches 1..311 of consensus"
18713..19133
repeat_region
/note="MER66-internal repeat: matches 4548..4919 of consensus"
complement(18872..19230)
/misc_feature
/note="match: GSS: Em:AQ005063"
19251..19719
/misc_feature
/note="match: GSS: Em:B14179"
19537..20290
repeat_region
/note="HERFH21 repeat: matches 4657..5784 of consensus"
20317..20382
repeat_region
/note="33 copies 2 mer ta 68 conserved"
20513..20666
repeat_region
/note="77 copies 2 mer tt 70 conserved"
20682..21008
repeat_region
/note="Alusg1 repeat: matches 1..306 of consensus"
21239..21553
repeat_region
/note="HURFS-P3 repeat: matches 4410..4713 of consensus"
21882..22254
repeat_region
/note="THRB repeat: matches 1..364 of consensus"
22302..22537
repeat_region
/note="MER66-internal repeat: matches 2186..2417 of consensus"
22538..22850
repeat_region
/note="AlusP repeat: matches 1..313 of consensus"
22851..23801
repeat_region
/note="MER66-internal repeat: matches 1210..2186 of consensus"
23905..23989
repeat_region
/note="MER66-internal repeat: matches 3017..3102 of consensus"

Alignment Scores:
Pred. No.: 1.4e-44 Length: 145880
Score: 623.00 Matches: 117
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 78.86% Indels: 0
DB: 9 Gaps: 0

US-09-997-610-2_COPY_1_149 (1-149) x HS302D9 (1-145880)
QY 33 ProGlyAlaProGlyLeuProGlyIntyrThrGlyGluIleSerGluMetThrLysCysPro 52
|||
Db 37262 CCTGTGCCCCAGGTTTACCAATATATACAGAGAGAAATTAAGTGAATGACAAAGGCC 37321
|||
QY 53 CysProAspIleGluArgSerAlaPheThrValLysLeuSerGlyLysLeuProLeuPro 72
|||
Db 37322 TGTCTGATATAGAAAGGTGACGCTTACTGTGAAGGTCAGTGAAGAACTCTCTTCT 37381
|||
```

```
QY 73 PheLysProIleIlePheThrGlyValLeuTyrAsnAlaGlnArgAspLeuLysGluAla 92
|||
Db 37382 TTCAAGCCCATATATCTTCACAGGGCTCTGTCAATGCCCAAGGATTTAAAGGAGGCC 37441
|||
QY 93 MetGlyValPheAlaCysArgValProGlyAsnTyrTyrSerSerPheAspValGluLeu 112
|||
Db 37442 ATGGAGTCTTGTTCGTCAGGGTGCCTGGCAATTACTACTCCAGCTTGTGAGCTG 37501
|||
QY 113 HisHisCysLysValAsnIleTrpLeuMetArgLysGlnIleLeuAlaAsnLysGluGlu 132
|||
Db 37502 CATCATTTGCAGAGTGAATATTGGCTTAATGAGAGCAAAATTTGGCTAATAGGAGNA 37561
|||
QY 133 IleSerLysGlnGlnSerIleGlnGluValIrrTrpValLeuLeuLysAla 149
|||
Db 37562 ATTCTTAAGCAGCAAGCAATTCAGAGGAGTGGTGGTGTGTAAAGCA 37612
|||

RESULT 2
TMSHP20A 1265 bp mRNA linear ROD 03-FEB-1999
LOCUS DEFINITION Tamias asiaticus mRNA for HP-20, complete cds.
ACCESSION D12974
VERSION D12974.1 GI:287467
KEYWORDS HP-20; collagen-like domain; hibernation-related protein; plasma protein.
SOURCE Tamias asiaticus liver cDNA to mRNA, clone:PCW20-7.
ORGANISM Tamias sibiricus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Sciurinae;
Tamias.
REFERENCE 1 (bases 1 to 1265)
AUTHORS Takamatsu N., Ohba K., Kondo J., Kondo N. and Shiba T.
TITLE Hibernation-associated gene regulation of plasma proteins with a
MEDLINE collagen-like domain in mammalian hibernators
REFERENCE 2 (bases 1 to 1265)
AUTHORS Takamatsu N.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1992) Nobuhiko Takamatsu, School of Hygienic
Sciences, Kitasato University, Laboratory of Molecular Biology;
1-15-1 Kitasato, Sagamihara, Kanagawa 228, Japan (Tel:0427-78-9408,
Fax:0427-78-9403)

FEATURES
Source Location/Qualifiers
1..1265
/organism="Tamias sibiricus"
/db_xref="taxon:64680"
/clone="PCW20-7"
/tissue="type="Liver"
89..679
/note="collagen-like domain at nt 161-277"
/codon_start=1
/product="HP-20"
/protein_id="BA02351.1"
/db_xref="GI:287468"
/translation="MTDVKRAIFVLAQNVLVNDVSCSGPPGPGYGVGVPPPPGPG
PGQPGAARGPDPGPKGSPVAKPCPKRSAPFVFKSGRIPLPSPSEVVFTEVLVYTRDL
KESTYGVNCPVGPVHFSFDVLYHKVKGITGLMNHIOVMEKHOLSKREVENAGAMI
MLPRQDKVLEADVETEEPPQARVIVYFSGLISS"
89..157
sig_peptide
158..676
mat_peptide
/product="HP-20"
/polya_site 1238..1243
BASE COUNT 354 a 304 c 288 g 319 t
ORIGIN

Alignment Scores:
Pred. No.: 1.34e-20 Length: 1265
Score: 330.00 Matches: 69
Percent Similarity: 61.22% Conservative: 21
Best Local Similarity: 46.94% Mismatches: 45
Query Match: 41.77% Indels: 12
```

[illegible][illegible]



Db 481 GTAACAGCATCCGGTCGCTTTTACTGTACTTCTTCCAAAGCTTACCAGCGGTA 540  
 QY 73 PheLysProIleIlePheThrClyValLeuTyrAsnIleGlnArgSpleuLysGluAla 92  
 Db 541 GGTGCTCCCATCCATTGATGATGATCTGTACAACAGCAGCAGCTGTGACCAAGA 600  
 QY 93 MetGlyValPheAlaCysArgValProGlyAsnTyrIleSerSerPheAspValGluLeu 112  
 Db 601 TCTGGAATCTTACTGTAAGATGATCCAGCATATATATTCTTCATACATCATGTG 660  
 QY 113 HIsHsICysLysValAsnIleTyrLeu-MetArgLysGlnIleLeuAlaAsnLysGlu 132  
 Db 661 AAGGGACTCAGCTTGGTGGTACGCTGTATAAGATGCA-----GCGCCACA 708  
 QY 132 uIleSerLysGlnIleSerIleGlnIleValPheTyrValLeuLys 148  
 Db 709 ATGTACACGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 757  
 RESULT 6  
 TMSHP27 1385 bp mRNA linear ROD 03-FEB-1999  
 LOCUS Tamias asiaticus mRNA for Hp-27, complete cds.  
 DEFINITION D12976  
 ACCESSION D12976  
 VERSION D12976.1 GI:287471  
 KEYWORDS HP-27; collagen-like domain; hibernation-related protein; plasma protein.  
 SOURCE Tamias asiaticus liver cDNA to mRNA, clone:PCM27-3.  
 ORGANISM Tamias sibiricus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae; Tamias.  
 REFERENCE 1 (bases 1 to 1385)  
 AUTHORS Takamatsu, N., Ohba, K., Kondo, J., Kondo, N. and Shiba, T.  
 TITLE Hibernation-associated gene regulation of plasma proteins with a collagen-like domain in mammalian hibernators  
 JOURNAL Mol. Cell. Biol. 13 (3), 1516-1521 (1993)  
 MEDLINE 93180798  
 REFERENCE 2 (bases 1 to 1385)  
 AUTHORS Takamatsu, N.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-AUG-1992) Nobuhiko Takamatsu, School of Hygienic Sciences, Kitasato University, Laboratory of Molecular Biology, 1-15-1 Kitasato, Sagamihara, Kanagawa 228, Japan (Tel:0427-78-9408, Fax:0427-78-9403)  
 FEATURES  
 source  
 1..1385  
 /organism="Tamias sibiricus"  
 /db\_xref="taxon:64680"  
 /clone="PCM27-3"  
 /tissue\_type="liver"  
 65..712  
 /note="collagen-like domain at nt 191-307"  
 /codon\_start=1  
 /product="Hp-27"  
 /protein\_id="BAA02353.1"  
 /db\_xref="GI:287472"  
 /translation="MYEAKRSPFMGAGIMIALSVLMHVVCSSETOGNPESENCVPG  
 OGPGMRGPGTGGKPGPGMNGPGLGPPGPPGVNVCNCHSGTSAFAVKNELPPA  
 PSQVYIEALHDAQHFDATGTCFTCPYGLIOPGFHIEAYQRAKVSLLMNGVOYA  
 EREAADQGYEHISGTAILOGMEDRWLEKNLSQDLEGRVQAVFSGLIHEN"  
 65..154  
 sig\_peptide  
 mat\_peptide  
 155..709  
 /product="Hp-27"  
 polyA\_signal  
 1369..1374  
 polyA\_site  
 1385  
 BASE COUNT 439 a 285 c 311 g 350 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 4,92e-06 Length: 1385  
 Score: 167.50 Matches: 51  
 Percent Similarity: 40.12% Conservative: 14  
 Best Local Similarity: 31.48% Mismatches: 68

Query Match: 21.20% Indels: 29  
 DB: 10 Gaps: 4  
 us-09-997-610-2\_copy\_1\_149 (1-149) x TMSHP27 (1-1385)  
 QY 6 ValLeuIleThrAlaValIleGluHisVal----- 15  
 Db 116 ATTTTGGCCCTGCTGTACTAATGACGCGTGCTGTGTGAAACTCAAGAAATCTGAA 175  
 QY 16 -----GluValAlaGlyPro-----Pro 21  
 Db 176 TCCGTATATGTTCCAGAGACTTCAAGSACTCCAGGATCGGGGTCGCCAGTACACT 235  
 QY 22 AlaHisProArgProProGluGluValGlyProProGlyAlaProGlyLeuProGlyTyr 41  
 Db 236 GGAATACAGAGCCACAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 295  
 QY 42 ThrGlyGluIleSerGluThrLysCysProCysProAspIleGluArgSerAlaPhe 61  
 Db 296 CCAGCC-----ATGACT---GTGAACCTCCACAGCAAGAAACATCGGCTTT 340  
 QY 62 ThrValLysLeuSerGlyLysLeuProLeuProPheLysProIleIlePheThrGlyVal 81  
 Db 341 GCACTGAAGCAATGACGTGCCCCCAGCTCCCTCCAGCCGCTGATCTTCAAGAAACC 400  
 QY 82 LeuTyrAsnIleGlnArgSpleuLysGluAlaMetGlyValPheAlaCysArgValPro 101  
 Db 401 CTGCATGACGCTCAGGACACTTGTATCTGCGACACTGCTGCTGCTGCTGCTGCTGCTG 460  
 QY 102 GlyAsnTyrTyrSerSerPheAspValGluLeuHIsHsICysLysValAsnIleTyrLeu 121  
 Db 461 GGACTCTACCAAGTTTGGATTTCATTGAACTGTCCACAGCGCTGTAAAGTACGCTC 520  
 QY 122 MetArgLysGlnIleLeuAlaAsnLysGluGluIleSerLysGlnIleSerIleGlnIle 141  
 Db 521 ATGGAATGCGACCCCAAGTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 580  
 QY 142 ValThr 143  
 Db 581 ATTTC 586  
 RESULT 7  
 CHKCX 2609 bp DNA linear VRT 28-APR-1993  
 LOCUS Chicken type X collagen gene.  
 DEFINITION M13496  
 ACCESSION M13496  
 VERSION M13496.1 GI:211699  
 KEYWORDS collagen.  
 SOURCE Chicken red blood cell DNA, clone pYN92E1; and embryo chondrocyte, cDNA to mRNA, clone pYN3116.  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
 REFERENCE 1 (bases 1 to 2609)  
 AUTHORS Nimni, Y., Gordon, M., van der Rest, M., Schmid, T., Linsemayer, T. and Olsen, B. R.  
 TITLE The developmentally regulated type X collagen gene contains a long open reading frame without introns  
 JOURNAL J. Biol. Chem. 261 (11), 5041-5050 (1986)  
 MEDLINE 86168227  
 PUBMED 3082876  
 COMMENT [1] has as yet not determined the start codon of the collagen X gene. The open reading frame starts at position 166. There are several 'atg' codons (starting at positions 190, 214, 256, 274 and 310) following the 'taa' at position 163.  
 FEATURES  
 source  
 1..2609  
 /organism="Gallus gallus"  
 /db\_xref="taxon:9031"  
 /cell\_type="chondrocyte"  
 intron  
 CDS  
 <380..2208



VERSION	AX332258.1	GI:18122892
KEYWORDS		
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	Young, P. E., Augustus, M., Carter, K. C., Ebner, R., Endres, G., Horrigan, S., Soppel, D. R. and Weaver, Z.	
TITLE	Cancer gene determination and therapeutic screening using signature gene sets	
JOURNAL	Patent: WO 0194629-A 2767 13-DEC-2001; Avalon Pharmaceuticals (US)	
FEATURES	Location/Qualifiers	
source	1..2235	
BASE COUNT	524 a 632 c 693 g 386 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	1.87e-05	2235
Score:	163.50	Matches: 42
Percent Similarity:	42.14%	Conservative: 17
Best Local Similarity:	30.00%	Mismatches: 48
Query Match:	20.70%	Indels: 33
DB:	6	Gaps: 5
US-09-997-610-2_COPY_1_149 (1-149) x AX332258 (1-2235)		
QY	19 GLYPProPProAlaHisProArgrProProGluGluValAlGlyProProGlyAlaProGlyLeu	38
Db	1660 GGCCAGCGCTGGCCCTTCAGACGCCAGGCCCTCCAGACCTCCAGACGCCAGCGCTGG	1719
QY	38 -----	38
Db	1720 ATGCCCCCTTACACACCACCCAGGAGAGATATGCCACATATGGCGCTGGGAATTGAT	1779
QY	39 -----ProIntyrThrGlyGluLeuSerGluMetThrIysCysProCysPro	54
Db	1780 GGGGTGAACCCCGCATGCTCAGGGG-----GCTAAGAAAGGCMAAGATGGAGGCCA	1833
QY	55 AsP1leG1uArGserAlaPheThrValIlysLeuSerGlyIysLeuProLeuProPheIys	74
Db	1834 GCCATATGAGTGGCTTCATTACCCCGACGCTAACCGCACCTTTCCACCGGTGGGGGC	1893
QY	75 ProIleIlePheThrGlyValLeuIyrAsnIaGlnArGaspLeuIysGluAlaMetGly	94
Db	1894 CCGATGAAGTTTAAACAATGCTGTATACGGCGACAGACCTACCAACCCGACAGACGGC	1953
QY	95 ValPheAlaCysArGValProGlyAsnIyrIserSerPheAspValGluLeuHisHis	114
Db	1954 ATCTTACCGTCGTAGAGTCCCTGTGTCTACACTTTTGCATATCCACGTT-----CAC	2004
QY	115 CysIys---ValAsnIleTrpLeuMetArGlySclnIleuAlaAsnIysGluGluIle	133
Db	2005 TGCAGGGGGGAGACCTGTGGGTT-----GCTCATTAACAAGACACAGACCGCCGTG	2055
RESULT 10		
LOCUS	HSCOL8A1	2235 bp mRNA linear PRI 04-AUG-1992
DEFINITION	Human COL8A1 mRNA for alpha 1(VIII) collagen.	
ACCESSION	X57527	
VERSION	X57527.1	GI:30081
KEYWORDS	collagen alpha 1 type III; extracellular matrix protein.	
SOURCE	Homo sapiens.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	Olsen, B. R.	
TITLE	Direct Submission	
JOURNAL	Submitted (05-FEB-1991) B. R. Olsen, Dept of Anatomy and, Cellular	

[illegible]

[illegible][illegible]

```

/clone_11b="NIH_MGC_68"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
238. .2472

```

BASE COUNT	624	a	698	c	753	g	431	t
ORIGIN								
Alignment Scores:								
Pred. No.:	2.6e-05	Length:		2506				
Score:	162.50	Matches:		42				
Percent Similarity:	42.75%	Conservative:		17				
Best Local Similarity:	30.43%	Mismatch:		50				
Query Match:	20.57%	Indels:		29				
DB:	9	Gaps:		5				
US-09-997-610-2_COPY_1_149 (1-149) x BC013581 (1-2506)								



QY	77	IIlethetHcIyValleuTyrAsnAlaGlnArGAspLeuLysGluAlaMetIyAlaPhe	96
		.....	.....
Db	72029	AACTGTAAACAACCTGGTGTATACGGCAGACAGAACACCCGACGACGACATCTTC	71970
Qy	97	AlAcTysArGValProGIIyAsnTyrTysSerCerPheAspAlaGluLeuHisHisGlySlys	116
		.....	.....
Db	71969	ACCTGTAGAGTCCCTCGTGTCTACTACTCTTTCATACACAGCTT	71919
		.....	.....
QY	117	---ValAsnIIleTyrPheuMetArGlySglIIleLeuAlaAsnLysGluIle	133
		.....	.....
Db	71918	GGGGCCACAGCTGTGGGTT	71874
		.....	.....
RESULT 14			
HS302D9/C			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			

repeat_region	/note="MR5A repeat: matches 26. .187 of consensus"
repeat_region	1033. .1336
repeat_region	/note="AluSp repeat: matches 1. .299 of consensus"
repeat_region	1450. .1583
repeat_region	/note="MR repeat: matches 24. .160 of consensus"
repeat_region	1687. .1752
repeat_region	/note="L2 repeat: matches 2593. .2661 of consensus"
repeat_region	2350. .2660
repeat_region	/note="AluSc repeat: matches 3. .309 of consensus"
repeat_region	2684. .2981
repeat_region	/note="AluSq repeat: matches 2. .300 of consensus"
repeat_region	3323. .3343
repeat_region	/note="MT1E repeat: matches 116. .136 of consensus"
repeat_region	3344. .3652
repeat_region	/note="AluY repeat: matches 1. .309 of consensus"
repeat_region	3653. .3928
repeat_region	/note="MT1E repeat: matches 136. .359 of consensus"
repeat_region	3929. .4278
repeat_region	/note="ME1B repeat: matches 3. .364 of consensus"
repeat_region	4279. .4485
repeat_region	/note="MT1E repeat: matches 359. .568 of consensus"
repeat_region	5073. .5176
repeat_region	/note="52 copies 2 mer ct 78 conserved"
repeat_region	5181. .5491
repeat_region	/note="AluJb repeat: matches 1. .311 of consensus"
repeat_region	6369. .6485
repeat_region	/note="L2 repeat: matches 2579. .2705 of consensus"
repeat_region	6647. .6685
repeat_region	/note="MADE1 repeat: matches 1. .23 of consensus"
repeat_region	6686. .6987
repeat_region	/note="AluXx repeat: matches 1. .302 of consensus"
repeat_region	6988. .7036
repeat_region	/note="MADE1 repeat: matches 23. .77 of consensus"
repeat_region	7482. .7754
repeat_region	/note="AluJb repeat: matches 9. .290 of consensus"
repeat_region	7775. .8060
repeat_region	/note="AluJo repeat: matches 1. .295 of consensus"
repeat_region	8414. .8511
repeat_region	/note="L2 repeat: matches 2553. .2706 of consensus"
repeat_region	8914. .9030
repeat_region	/note="MR repeat: matches 147. .262 of consensus"
repeat_region	9110. .9280
repeat_region	/note="MR repeat: matches 91. .262 of consensus"
repeat_region	9283. .9412
repeat_region	/note="MR repeat: matches 15. .144 of consensus"
repeat_region	9521. .9679
repeat_region	/note="FAM repeat: matches 3. .161 of consensus"
repeat_region	9820. .10225
repeat_region	/note="MRB repeat: matches 2. .425 of consensus"
repeat_region	complement(10179. .10678)
repeat_region	/note="match: GSS: Em: B56592"
repeat_region	complement(10204. .10728)
repeat_region	/note="match: GSS: Em: A0701486"
repeat_region	complement(10249. .10706)
repeat_region	/note="match: GSS: Em: A0225495"
repeat_region	10312. .10383
repeat_region	/note="MR repeat: matches 79. .150 of consensus"
repeat_region	10718. .11310
repeat_region	/note="match: GSS: Em: B14024"
repeat_region	10784. .11201
repeat_region	/note="match: GSS: Em: B43566"
repeat_region	11838. .11946
repeat_region	/note="MR repeat: matches 20. .137 of consensus"
repeat_region	12174. .12445
repeat_region	/note="L2 repeat: matches 1988. .2275 of consensus"
repeat_region	12444. .12642
repeat_region	/note="MR repeat: matches 63. .241 of consensus"
repeat_region	13017. .13369
repeat_region	/note="match: STR: Em: C49301"
repeat_region	13331. .13397
repeat_region	/note="MR repeat: matches 174. .244 of consensus"
repeat_region	13398. .13698
repeat_region	/note="AluSp repeat: matches 1. .302 of consensus"

```

repeat_region 13699. .13810
/note="MIR repeat: matches 76. .174 of consensus"
repeat_region 13806. .13919
/note="MIR repeat: matches 77. .189 of consensus"
repeat_region 13945. .14060
/note="MIR repeat: matches 24. .142 of consensus"
repeat_region 14061. .14367
/note="Alu repeat: matches 1. .301 of consensus"
repeat_region 14368. .14452
/note="MIR repeat: matches 141. .225 of consensus"
repeat_region 14589. .14679
/note="MIR repeat: matches 173. .262 of consensus"
misc_feature 14597. .15201
/note="match: GSS: Em:A053482"
14616. .15060
/note="match: GSS: Em:A0370601"
14868. .15040
/note="MIR repeat: matches 49. .233 of consensus"
repeat_region 15071. .15188
/note="L2 repeat: matches 2112. .2239 of consensus"
repeat_region 15304. .15399
/note="MLTIB repeat: matches 1. .99 of consensus"
repeat_region 15450. .15662
/note="AluSg1 repeat: matches 2. .114 of consensus"
repeat_region 15669. .15727
/note="MLTIB repeat: matches 119. .178 of consensus"
repeat_region 15728. .16027
/note="AluSc repeat: matches 1. .299 of consensus"
repeat_region 16028. .16245
/note="MLTIB repeat: matches 178. .390 of consensus"
16546. .16854
/note="Alu repeat: matches 1. .300 of consensus"
repeat_region 18296. .18323
/note="MSTR repeat: matches 2. .29 of consensus"
18324. .18392
/note="MER6-internal repeat: matches 4919. .4993 of consensus"
repeat_region 18393. .18712
/note="AluDb repeat: matches 1. .311 of consensus"
18713. .19133
/note="MER6-internal repeat: matches 4548. .4919 of consensus"
misc_feature complement(18872. .19230)
/note="match: GSS: Em:AQ005063"
19251. .19719
/note="match: GSS: Em:BI4179"
19537. .20290
/note="HERFH21 repeat: matches 4657. .5784 of consensus"
repeat_region 20317. .20382
/note="33 copies 2 mer ta 68 conserved"
20513. .20666
/note="77 copies 2 mer tt 70 conserved"
20682. .21008
/note="AluSg1 repeat: matches 1. .306 of consensus"
21239. .21553
/note="HUES-P3 repeat: matches 4410. .4713 of consensus"
21882. .22254
/note="THB1 repeat: matches 1. .364 of consensus"
22302. .22537
/note="MER6-internal repeat: matches 2186. .2417 of consensus"
22538. .22850
/note="AluSp repeat: matches 1. .313 of consensus"
22851. .23801
/note="MER6-internal repeat: matches 1210. .2186 of consensus"
23905. .23989
/note="MER6-internal repeat: matches 3017. .3102 of

```

Alignment Scores:  
Pred. No.:  
Score:  
Percent Similarity:

0.00201 Length: 145880  
162.50 Matches: 42  
50.89% Conservative: 15

```

Best Local Similarity: 37.50% Mismatches: 46
Query Match: 20.57% Indels: 9
DB: 9 Gaps: 3
US-09-997-610-2_COPY_1_149 (1-149) x HS302D9 (1-145880)
QY 37 G1LEUPROGINTYRTHRG1GLIUESERGIMETHTHRCYSPROCSYPROSP5P1E 56
||||| :||| :||| :||| :|||
DB 90955 GGAATCCCGTTCGTTCTATGACAAA-----CAAAGACCAAGACACTAGAACAGATATT 90900
QY 57 Glu-----ArgSerAlaPheThrValLysLeuSerGlyLysLeuPro 70
||| ||||| ||||| ||||| ||| |||||
DB 90899 GAAAAAGCTTCATCCCAACCAAGATCTGATTTGGTGTAACATGATGACCCCTCCA 90840
QY 71 LEUPROPHELYSPTIIEIIEPHEHTRG1VALLEUTYASNAIIGLIRGASPLEULYS 90
:||||| :||||| :|||
DB 90839 GTGCCCTTCTACCCATCTCTTCAAGAGAACCCCTTAAATACATCAGTTCATTTCCAT 90780
QY 91 G1UALAMEGL1VALPHEALACYSARVALP1ROGL1ASNTYTRFSERSEPHASVAL 110
||| ||| ||| ||| ||| ||| ||| |||
DB 90779 TTCTCTGGGAAGGATTCACCTGTACCATCCCTAGTGTTACCATTTGTTGAATTT 90720
QY 111 G1LEUHLISHSCYSLYSVALSNILETRPLEUMETATRG1G1LE--LEUALASN 129
||||| ||| :||| :||| ||| ||| |||
DB 90719 GAGCTGTTTCACAGAGCTGTGCACAAATGCGGTATTATGAGAAATGATCATATGAGAAAT 90660
QY 130 LYSGLIUGLIIEISERLYSG1NG1INSERIEG1NG1U 141
||| :||| ||| ||| |||
DB 90659 AAACATGCTGTGGCCAGAGATGCTATGACGAGAA 90624
RESULT 15
AC120598
LOCUS 152354 bp DNA linear HTG 23-JUL-2002
DEFINITION Homo sapiens, *** SEQUENCING IN PROGRESS ***, 33 unordered pieces.
ACCESSION AC120598
VERSION AC120598.3 GI:21902634
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 152354)
REFERENCE
Albrooks,S.L., Amaralunge,H.C., Adio-Oduolu,A.B., Ali-osman,F.R., Allen,C.,
Barbarta,J., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Bunhay,C., Burch,P., Burkelt,C., Butrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chan,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Dayla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
DeJaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinn,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhardt,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,M., Gunarrane,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsi,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudh,S.,
Karlsom,E., Kelly,S., Khan,U., King,L., Korvab,J., Kovar,C.,
Kratovic,J., Kuresh,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtege,O., Lieu,C., Liu,C., Liu,W., Louised,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mahoney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokwenkwo,S., Ogun,M., Okunodu,G.,
Oraunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojuboan,I., Rolfe,M., Rutz,S., Savery,G.,

```



```
Db 84256 GAGATGCTGCATTACCGCCGAGCTAACCGACCTTCCACCGGTGGGGCCCCAGTG 84315
Oy 77 IlePheThrGlyValLeuTyrAsnAlaGlnArgAspLeuLysGluAlaMetGlyValPhe 96
    ||| ::||| ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 84316 AAGTTTAAACAACCTGCTGTATACGGCAGACAGAACTACACCCGCGAGACAGGCATCTTC 84375
Oy 97 AlaCysArgValProGlyAsnTyrTyrSerSerPheAspValGluLeuHisCysLys 116
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84376 ACCTGTGAGGTCCCTGTGTCTACTACTTTCATACACCGTT-----CACTGCANAG 84426
Oy 117 ---ValAsnIleTyrLeuMetArgLysGlnIleLeuAlaAsnLysGluGluIle 133
    ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 84427 GGGGCAACGCTGGGTT-----GCTCTATTCAGACACAGACAGCCCGTG 84471
```

Search completed: February 19, 2003, 23:04:03  
Job time : 2208.08 secs

**THIS PAGE BLANK (USPTO)**



## JOURNAL COMMENT

Unpublished (2000)  
 Contact: xingwu lu, lianxian cui, yonghai li  
 Department of Biochemistry  
 Institute of Basic Medical Science, Peking Union Medical College  
 Dongdan Sanliao 5, Beijing, P.R.C, 100005  
 Tel: 86-010-65296951  
 Email: luxingwu@263.net  
 full-length and coding sequence.

## FEATURES

Location/Qualifiers

1..2615  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone-lib="human bone marrow cDNA library"  
 /tissue-type="bone marrow"  
 /note="Organ: uterus; Vector: pOT7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 705 a 531 c 641 g 738 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 1,23e-96 Length: 2615  
 Score: 1184.50 Matches: 244  
 Percent Similarity: 74.65% Conservative: 21  
 Best Local Similarity: 68.73% Mismatches: 76  
 Query Match: 49.03% Indels: 15  
 DB: 10 Gaps: 3

US-09-997-610-2\_COPY\_18\_459 (1-442) x BE420422 (1-2615)

QY 101 AsnIleTfPleuMetArgLysGlnIleLeuAlaAsnLysGlnIleSerLysGlnIle 120  
 Db 494 AACCTTGAGCTTGAGAGATCATTTAGGGTATCTGGCAGAAATTTCTAAGCAGCA 553  
 QY 121 SerIleGlnIleValThrTrpValLeuLeuValAlaPheSerPheIleArgIleVal 140  
 Db 554 AGCATTCAGAGGTRACTGGGTCTCTTTAAAGGCATTTTAAAGGAGCAGG 613  
 QY 141 HisLysSerSerGluAsnLeuHisProAspAsnValIleLysLysLysAsnProPheSer 160  
 Db 614 CATGAAGTTGGAAATTTGCAGCTGCAGCAATGACAAAGAAATCCCATTTCT 673  
 QY 161 GluGlyLysPheLysLeuAlaIleGlyIleCysIleCysAsnIleLysValAsn 180  
 Db 674 GAGGAGAAATCAAGCTGGCTGCAGAAATTTGCATATGTAACAGAGTTAAATGTTAAT 733  
 QY 181 ProGlnAspAsnGlyGluAsnIleSerTrpThrCysGlnArgSerSerGlnIleSerIle 200  
 Db 734 CCCCAAGCATGGGGAAATGCTCCAGGGCATGCCAGAGGCTTCACAGCAGCCCTC 793  
 QY 201 LysSerLeuAlaTrpArgProArgArgLysTrp---PheCysGlyThrGlyProGlySer 219  
 Db 794 CCATCAGAGCCTGCAGGCTAGAGAGAAAGATCTCTGGGCCAGGCCAGGGTCC 853  
 QY 220 LeuGlyCysValGlnProArgAspLeuValProCysValProValAsnSerAla----- 237  
 Db 854 CCATCTGTGTGCAGCTGGGGACTGTGGCTGTGTCCACCTCTCCAGCCATGGCT 913  
 QY 238 -----ValAlaSerGlnIleAlaSerProLysProLysProTrp 248  
 Db 914 GAAAGGGCCAAATAGAGCTCGGATTTGGCTTCAGAGGGTCAAGCTGTAAAGCCTTGG 973  
 QY 249 GlnLeuProSerGlyValGluProValGlyAlaLysLysSerArgIleGluValTrpGlu 268  
 Db 974 CAGCTTCATGTGTGTGAGCTGCAGGTCACAGCAAGTCAAAATTTGGGGTTGGGAA 1033  
 QY 269 ProProlIleArgPheGlnLysIleTyrGlyAsnProTrpMetProArgGlnLysPheAla 288  
 Db 1034 CCTGTGCTAGATTTCAGAAAGATGTATGAAATGCCGTGATGCCCTGCAGAGATCTGCT 1093

QY 289 ValGlyValIleSerSerTrpArgThrSerAlaArgValGlnLysGlyAsnValGly 308  
 Db 1094 GTAGGGAGAGGGGCCCTCATGAGATCTCTCTAGGGCAGTACAGAAAGAAATGTGGGA 1153  
 QY 309 TrpGluProProHisArgValProSerGlyAlaProSerSerArgAlaValArgArgSer 328  
 Db 1154 TCGAGACCCCAACAGATCCCTACTGTGGGCGCCCTGTGAGAGTGTGAGAAAGAGG 1213  
 QY 329 ProProSerSerArgLeuGlnLysGlyArgSerThrAspSerLeuGlnHisValProGlu 348  
 Db 1214 CCACCATCTCCAGACCCCAAGATGATGATCCATGACAGCTTGTCTCCATGACCTTGA 1273  
 QY 349 LysSerThrAspThrGlnCysGlnProValLysAlaAlaGlyMetGlnSerValProTyr 368  
 Db 1274 AAACCCGACAGC---ACAC-CAGCCCATGAAACACACTAGGACGAGGCTGTACCTGC 1329  
 QY 369 LysThrValAlaGlnLeuThrLysThrValGlyIleTyrLeuLeuHisCysHisAsp 388  
 Db 1330 AAGCCACAGGGCGGAGCTGCCCAAGATCATGGGAACCCACCTCTGTGATCAGCATGAC 1389  
 QY 389 LeuAspValArgHisGlyValLysArgAspHisPheGlyValAlaLeuArgPheAspCysPro 408  
 Db 1390 CTGGATGTGAGATTGTGAGATTGAAGAGATCATTTTGAGCTTTAAGATTGTGACTGCC 1449  
 QY 409 ThrGlyPheArgThrTyrMetGlyProVal-ProLeuCysPheGlyGlnPhePhePro 428  
 Db 1450 TCTAGATTTCAGACTTGATGGAGCTGTAGCCCTTGTGTTACTCATATTCTCCATT 1509  
 QY 428 eGlyThrAlaValPheThrGlnCysLeuTyrLeuHisCysMet 442  
 Db 1510 TGGACAGCTGTGTTTACCAATACCTGTACCCCATTTGTATC 1552

## RESULT 2

BC004496 2009 bp mRNA linear HTC 12-JUL-2001  
 LOCUS  
 DEFINITION Homo sapiens, similar to hypothetical protein FLJ14036, clone IMAGE:383133, mRNA.

ACCESSION BC004496  
 VERSION BC004496.1 GI:14709139  
 KEYWORDS HTC.

SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.

REFERENCE 1 (bases 1 to 2009)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK COMMENT

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbs-remail.ncl.nih.gov](mailto:cgapbs-remail.ncl.nih.gov)  
 Tissue Procurement: ATCC/DCTD/DRP  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
 contact: [amadansystemsbiology.org](mailto:amadansystemsbiology.org)  
 Anup Madan, Rachel Dickhoff, Jessica Fahy, Stephanie Ford, Julia Greene, Mark Ketteman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRML Plate: 14 Row: d Column: 7  
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis

This clone has the following problem: frame shifted.

## FEATURES

source

1..2009

Location/Qualifiers



Db	61	TTGTAACAGAGAGTTAAATGTTATATCCCAACAAATGGGAAAAATGTCCTCAGGCGATCC	120
Qy	193	GLnargSerSerGlnInSerILeLysSerLeuAlaTrpArgProArgArgIysTrp---	211
Db	121	CAGAGGCTCTTCAAGACGCGCCCTCCCATCAAGGCGCTGCAGGCGCTGAGAGAAAGTA	180
Qy	212	PheCysGlyThnGlyProGlySerLeuGlyGlyValGlnProArgAspLeuValProCys	231
Db	181	TTCTCTTGCGCAAGCCCAAGGTCGCCATGCTGTGTGCGAGCCTGGGAGCTTGGGCTTGT	240
Qy	232	ValProValAsnSerAla-----ValAlaSer	240
Db	241	GTCCACAGCTCTCCACCATGGCTGAAGGGGCCAACATGAGCTGGAGATTGGCGTTCA	300
Qy	241	GLnGlyAlaSerProLysProTrpGlnLeuProSerGlyValGlnProValGlyAlaLys	260
Db	301	GAGGCTGCAGAGCTGTAAAGCTTGGAGCGTTCCATGCTGGTGTGGCTGCAGTGCACAG	360
Qy	261	LysSerArgIleGlnValTrpGlnProLleArgPheGlnLysIleArgLysAsnPro	280
Db	361	AAGCAAAAATTGGGTTTGGGAACCTGCGCTGATTTGGAAGATGTGGAATGGC	420
Qy	281	TrpMetProArgGlnLysPheAlaValGlyValGlySerSerTrpArgTrsAlaArg	300
Db	421	TGGATGCGCCCTGGAGAGAGTCTGCTTAGGGGAGGGGCCCTCATGAGAAATCTCTGCTAG	480
Qy	301	ValValGlnLysArgLysAsnValGlyTrpGlnProProHisArgValProSerGlyAlaPro	320
Db	481	GCATACACAGAGGAAATGTGGATCGAGAGCCCAACAGACAAATCCTACTGGGCGACCG	540
Qy	321	SerSerArgAlaValArgArgSerProProSerSerArgLeuGlnLysGlyArgSerThr	340
Db	541	CCTAGTGCAGCTGTGAGAAGAGGCGACACATCTCCAGACCCAGAAATGATATGATCCACT	600
Qy	341	AspSerLeuGlnHisValProGlnLysSerThrAspTrpGlnCysGlnProValLysAla	360
Db	601	GACAGCTGCTGTCATCCACTGGAAGAAACCGAGAC--AACAC-CAGCCCAAGAAAGCA	656
Qy	361	AlaGlyMetGlnSerValProTrpLysThrValAlaValAlaGlnLeuThrLysThrValGly	380
Db	657	GCTAGGCAAGAGGCTTACCTCGCAAGACGACAGAGGGCGAGCGTCCCAAGATCATGGGA	716
Qy	381	ILeTrpLeuLeuHisCysHisAspLeuAspValArgHisGlyValLysAsnAspHisPhe	400
Db	717	ACCACACTCTTCATCAGCATGACATGACCTGGATGTGAGATTGTGAGTAAAGAGATCACTTTT	776
Qy	401	GlyAlaLeuArgPheAspCysProThrGlnLysPheArgTrpTrpMetGlyProVal-Prole	420
Db	777	GGAGCTTAAAGATTTACTGCCCCCTCTGATTTTACGACTTCATGGGGCTGTAGGCCCT	836
Qy	420	uCysPheGlyGlnPhePheProPheGlyThrAlaValPheThrGlnCysLeuTrpLeuHI	440
Db	837	TTGTTTACTCATTTCTCCCATTTGGAACAGCTGTGTTTACACATAACTGTACCCCA	896
Qy	440	sCysMet 442	
Db	897	TTGTATC 903	
RESULT 4	BM457166	898 bp	mRNA
LOCUS	BM457166		linear
DEFINITION	AGENCOURT_6411690 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5583427		EST 05-FEB-2002
ACCESSION	BM457166		
VERSION	BM457166.1		GI:18506206
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Ekakayota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	NIH-MGC <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a> .		

TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA sequencing by: Amersham Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov Plate: LLM12346 row: h column: 20 High quality sequence start: 9 High quality sequence stop: 713. Location/Qualifiers				
FEATURES	1..898 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5583427" /clone_id="NIH_MGC_92" /tissue_type="embryonal carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: testis; Vector: pCMV.SPORT6; Site:1; NOT1; Site:2; Salt: Cloned unidirectionally; oligo-dt primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies Note: this is a NIH_MGC Library."				
BASE COUNT	225 a	228 c	254 g	191 t	
ORIGIN					
Alignment Scores:					
Pred. No.:	1.le-71	Length:	898		
Score:	902.50	Matches:	188		
Percent Similarity:	70.83%	Conservative:	16		
Best Local Similarity:	65.28%	Mismatches:	64		
Query Match:	37.36%	Indels:	21		
DB:	13	Gaps:	3		
US-09-997-610-2_COPY_18_459 (1-442) x BM457166 (1-898)					
Oy	171	CysilecysannglucsluileuansynalasnProdlinspansnglunsnilesertp	190		
Db	35	TGCTATAGCAGCAGAGAGCGCTATGTTATGCCAAGACATGGGAAATGTCACAG	94		
Oy	191	ThrcyslnargserseinglnserllelysserleualatrpargProargr---	209		
Db	95	TCATGTCAGAGACCTTCATGCAGCCCTCCATCACAGGCCCGACGCGCCAGAGGAA	154		
Oy	210	lystrphecysglyrhncglyprodysereuocysyalslgnprograspneual	229		
Db	155	AAGGGTTTTGTGGCTGGGGCCCAAGGCTCTGTGCTGTGTGACGCTTAGGGACTTGGTG	214		
Oy	230	ProctysvalProvalanser-----	238		
Db	215	CCCTGTGTCACACCTCTCTCACCATGTGGCTGAAGGACCAATAGAGCTTGGGCTGTG	274		
Oy	239	AlaserciluclyalaserProlyserProtrpelnleuProsercilyalglunProvalgly	258		
Db	275	GCTTCACAGAGGTGAGAGCCCAAGACATTGGACAGCTTCATGATGATGATGAGCTGTGGGT	334		
Oy	259	AlatyslysserargtleglualvaltrpblunProprollaargphecglmslleygly	278		
Db	335	GCAGAGATGTCACAGATTAGATTGGGAACCTCTGCGCTGATTTACAGATATATATGA	394		
Oy	279	AsndProtrpncProarglinslysphealavalglyalyserserttrpargThrsr	298		
Db	395	AACACCTCGATGTCACGCGCAAAAGTTTCTCTCACAGGCGCAAGCCCTCATGCGGAACCTCT	454		
Oy	299	AlargvalvalcglmslyslasynalglytrpblunProprohslsryalProsercily	318		
Db	455	CTTAGGGCACTGTGGGAAGGAATGTGGGGA-GAGCCCCACACAGATCCCTCTTGGG	513		
Oy	319	AlaroserseirrgalavalargargserProproser-----	331		

Db 514 GCACTGCTAGTAGGAGCTGTGAGAAAGCCACAGCTCCAGAACGGTAGAGACCCAGAAAT 573  
Qy 332 SerArgLeuGlnLysGlyArgSerThrAspSerLeuGlnHisValProGluLysSerThr 351  
Db 574 GGTAGACCCCGAGAAATGGATGATCAACACCACTTCATCTGTGCTGAGAAAGCTTCA 633  
Qy 352 AspThrGlnCysGlnProValLysAlaIleGluMetGlnSerValProTyrLysThrVal 371  
Db 634 GACACTCAACATCAGACCTGAGAAAGCCAGAGAGGGAGCTGTACTGTCAAAACCCACA 693  
Qy 372 ValAlaGluLeuThrLysThrValGlyLeuLysLeuHisCysHisAspLeuAspVal 391  
Db 694 GGGGGGAGCTGACACAAAGACCATGAGACCTATCTCTGATCAGATCAGACCTGATGATGG 753  
Qy 392 ArgHisGlyValLysArgAspHisPheGlyAlaLeuArgPheAspCysProThrGlyPhe 411  
Db 754 AGACATGGAGTCMAAGGATCATATGAGCTTAAATTTGACTGCCGCCACTGATTTT 813  
Qy 412 ArgThrTyrMetGlyProVal-ProLeuCysPheGlyGlnPhePheProPheGlyThrAl 431  
Db 814 CAGACTTCATGAGACCTGTAAACCCCTTGTGTTGGCCAAATTTCTCTCTTGGAGACAGC 873  
Qy 431 AlaPheThrGlnCysLeuTyr 438  
Db 874 TGTATTATCCCAATACCTGTAC 895  
RESULT 5  
B0893011  
LOCUS B0893011 937 bp mRNA linear EST 16-AUG-2002  
DEFINITION AGNCOURT\_8122304 Lupski\_dorsal\_root\_ganglion Homo sapiens cDNA  
clone IMAGE:6179271 5', mRNA sequence.  
ACCESSION B0893011  
VERSION B0893011.1 GI:22285025  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 937)  
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.  
J11LE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Dr. James R. Lupski  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LMN at:  
<http://image.llnl.gov>  
Plate: LLM13560 row: C column: 16  
High quality sequence stop: 722.  
Location/Qualifiers  
1. 937  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6179271"  
/clone\_lib="Lupski\_dorsal\_root\_ganglion"  
/sex="male"  
/tissue\_type="dorsal root ganglia"  
/dev\_stage="adult, 36 yr"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6 (Life Technologies); Site\_1:  
NotI; Site\_2: SalI; cDNA made by oligo-dT priming.  
Directionally cloned using the following adaptors:  
5'-TCGACCAAGCGGCGG-3' and  
5'-GACTAGTTCTAGATCGGAGCGGCCCTT(15)-3'. Size selected >  
1 kb for average insert length 1.7 kb. This is a primary  
library, non-amplified. Library constructed by Life  
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor  
College of Medicine) and is available through Life

BASE COUNT 235 a 226 c 258 g 214 t 4 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 2,46e-69 937  
Score: 877.00 Matches: 194  
Percent Similarity: 70.53% Conservative: 19  
Best Local Similarity: 64.24% Mismatches: 64  
Query Match: 36.30% Indels: 28  
DB: 14 Gaps: 4  
US-09-997-610-2\_copy\_18\_459 (1-442) x B0893011 (1-937)  
Qy 156 LysAsnProPheSerGluGlyLysPheLysLeuAlaIleGluCysIleCysAsnGlu 175  
Db 6 AAAATTGCAATTTTCGAATAGAAATTCAAGCCAGCTGCAGAAATTTGATAGGAAGAG 65  
Qy 176 GluLeuAsnValAsnProGlnAspAsnGlyGluAsnIleSerTrpThrCysGlnArgSer 195  
Db 66 GAGCCAAATGTTAAATCCCAAGACAAATGGGAAATGTCCTCAGGCCATGTACAGGCT 125  
Qy 196 SerGlnGlnSerIleLysSerLeuAlaTrpArgProArgArg---LysTrpPheCysGly 214  
Db 126 TCAGCGCAGTCCCTCCATCAGGCCAGAGCCAGAGGAGGAAAGTGTTCATGGG 185  
Qy 215 ThrGlyProGlySerLeuCysCysValGlnProArgAspLeuValProCysValProVal 234  
Db 186 CTGGGCCAGGGGTCCCAAGCATATGTGCAGCTGG-GGTTGTGCACATGCCATCCAGCT 244  
Qy 235 AsnSerAlaValAlaIleSer-----GluGlyAla 243  
Db 245 GCCCAGCTGTGGCTGMAAGGGGACAAATGTAGAGCTCGGGCCCTGGCTTTAGATGTGCA 304  
Qy 244 SerProLysProTrp-----GlnLeuProSerGlyValGlnProValGlyAlaLysLys 261  
Db 305 AGCCCATGCTCTGTGGTGTGTTGATGCTTCATGTGGTTGACCCGACAGGTGCACAGAA 364  
Qy 262 SerArgIleGluValTrpGlnProProIleArgPheGlnLysIleTyrGlyAsnProTrp 281  
Db 365 TCMAAATTTGAGCTTTTGGAACTCCACCTCAATTTCAAGAAATGTATGAAATCCCTGG 424  
Qy 282 MetProArgGlnLysPheAlaValGlyValGlySerSerTrpArgThrSerAlaArgVal 301  
Db 425 ATGCCAGGCAAAATTTTGCTCAAGGGTGGGGCTCTCATGGAGAACCTCTCTTGGGCA 484  
Qy 302 ValGlnLysGlyAsnValGlyTrpLeuProProHisArgValProSerGlyAlaProSer 321  
Db 485 GTGCAGAAAGGAAATGTGGGTCGAGTCCCAACAGAGTCCCTACTGGGCA----- 538  
Qy 322 SerArgAlaValArgArgSerProProSerSerArgLeuGlnLysGlyArgSerThrAsp 341  
Db 539 -----CTGTCTCCCAAGCC-CAGAAAGTAGATCCACTGAC 573  
Qy 342 SerLeuGlnHisValProGluLysSerThrAspTrpGlnCysGlnProValLysAlaIle 361  
Db 574 ACCTTGACACTGTGTGCTGGAAAGACACAGACACTCAATGCACTATGAAAGCAAC 633  
Qy 362 GlyMetGlnSerValProTyrLysThrValValAlaGluLeuThrLysThrValGlyIle 381  
Db 634 AGGAGGAGAGACTACCCCTGCAGAACACAGAGGCGCAGAGCTCTCAAGACSCATGGGAAC 693  
Qy 382 TyrLeuLeuHisCysHisAspLeuAspValArgHisGlyValLysArgAspHisPheGly 401  
Db 694 TACCTCTGTATACACATCAGCTGATGATGAGACATGTGAGTCAAGAGAACATTTTGA 753  
Qy 402 AlaLeuArgPheAspCysProThrGlyPheArgTrpTyrMetGlyProVal-ProLeu 421  
Db 754 GCTTAAGATTTTGACTGCCCTGTGATTTTGGACTGTGCAAGCCGTGTAGCCCTTTG 813  
Qy 421 SerPheGlnPhePheProPheGlyThrAlaValPheThrGlnCysLeuTyrLeuHisCys 441  
Db 814 ATTGGCCAGATTCTCCCAATTTGGAATGCGATATTATTC-CATGCSATATACSSCCATTTG 872

QY 441 smet 442  
|:::  
Db 873 TATC 876

RESULT 6  
BM471183  
LOCUS  
DEFINITION BM471183 1036 bp mRNA linear EST 05-FEB-2002  
5', mRNA sequence.  
ACCESSION BM471183  
VERSION BM471183.1 GI:18520225  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 1036)  
NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: c9apbs-r@mail.nih.gov  
Tissue Procurement: ATCC/DC/DTP  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLM12233 row: g column: 18  
High quality sequence stop: 661.  
Location/Qualifiers  
1. 1036  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5563049"  
/clone\_lib="NIH\_MGC\_72"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pCMV-Sport6; Site:1: NotI;  
Site:2: SalI; Cloned unidirectionally. Primer: oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."

BASE COUNT 254 a 239 c 294 g 247 t 2 others  
ORIGIN

Alignment Scores:  
Pred. No.: 1.38e-68 Length: 1036  
Score: 869.50 Matches: 177  
Percent Similarity: 73.09% Conservative: 24  
Best Local Similarity: 64.36% Mismatches: 60  
Query Match: 35.99% Indels: 15  
DB: 13 Gaps: 3

US-09-997-610-2\_COPY\_18\_459 (1-442) x BM471183 (1-1036)

QY 178 AsnValasnProGlnAspAsnGlyGluAsnIleSerThrCysGlnArgSerGln 197  
|:::  
Db 33 AATGTTATCCCAAGACATGGGAAATGCTCCAGTTCCGTCAGAGACCTTCACG 92

QY 198 GlnSerIleIysSerLeuAlaTrpArgProArgArg---LysTrpPheCysGlyThrGly 216  
|:::  
Db 93 CAGACCCCTCCATCAGAGCCAGAGGCCAGAGGAAAGAGGTGTTGTGGCTGGGT 152

QY 217 ProGlnSerLeuCysValGlnProArgAspLeuValProCysValPro----- 233  
|:::  
Db 153 TCAGGGTCCCATCTGTGTGAAGCCTAGGACTGTCCTGTCCTCCCTCCTCTCCA 212

QY 234 -----ValAsnSerAlaValAlaSerGluAlaSerPro 245  
|:::  
Db 213 GCCTTGCTGAAGGGCTAATGATATAGCTGTGCTGTGGCTTTAGAGGGGTGAAGCCCC 272

QY 246 LysProTrpGln---LeuProSerGlyValGlnProValGlnAlaLysSerArgIle 264  
|:::  
Db 273 AAGCCTTGGACACTTCTTCATGCTGTGTGAGCTGTGGGTGACAGAGTCAAGAACT 332

QY 265 GluValTrpGlnProProIleArgPheGlnLysIleTyrGlyAsnProTrpMetProArg 284  
|:::  
Db 333 GAGCTTTGGGAACCTCT--GCTAGATTCCAGAAAGATGATGTAATGCTGATGCCAGG 391

QY 285 GlnLysPheAlaValGlnLysSerSerTrpArgTrpSerAlaArgValGlnLys 304  
|:::  
Db 392 CAAAGTTTGCTGCAAGGTTGGGGCCCTCATGAGAAACCTCTAGAGGAGTGTGGAG 451

QY 305 GlyAsnValGlnTrpGlnProProIleArgValProSerGlyAlaProSerSerArgAla 324  
|:::  
Db 452 GGAATATGTGGGTGGAGGCCGCCAGAGTCCCTGCTGGGCACTGCTAGTGGAGCT 511

QY 325 ValArgArgSerProProSerSerArgLeuGlnLysGlyArgSerTrpAspSerLeuGln 344  
|:::  
Db 512 GTGAGAAAGAGGGCCACACATCCCGACAGCCAGAAATGTGATCATCAGACCTTGGAC 571

QY 345 HisValProGlnLysSerThrAspThrGlnCysGlnProValLysAlaAlaGlyMetGlu 364  
|:::  
Db 572 TGTGCACTTGAAAGAGCCGACAGACACTGACACACAGCCCATGAAAGCAGCTGGAG 631

QY 365 SerValProTrpTrpLeuValValAlaGluLeuThrLysTrpValGlyIleTyrLeuLeu 384  
|:::  
Db 632 GCTATACCTCGTGAAGACACAGGAGGAGAGTCCCTTAAGACCATGAAACCCACTTGTG 691

QY 385 HisCysHisAspLeuAspValArgHisGlyValLysArgAspHisPheGlyAlaLeuArg 404  
|:::  
Db 692 CATCAGACATGACCTGTGATATGAGACTTGAGACTCAAGAGATCATCTTTGGAGCTTGAA 751

QY 405 PheAspCysProThrGlyPheArgThrTyrMetGlyProVal-ProLeuGlySerGly 424  
|:::  
Db 752 TTGTGATTTGCCCGCGATGATTCAGACTGTGTGANGCCTGATAAACCTTTGTTGGGCA 811

QY 424 nPhePheProPheGlyThrAlaValPheThrGlnCysLeuTyr 438  
|:::  
Db 812 ATTTCTCCCATTTTGAATGGCTACATTTTACCAATACCTGTAC 854

RESULT 7  
BM472108 1050 bp mRNA linear EST 05-FEB-2002  
LOCUS AGENCOURT\_6465359 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:5539381  
DEFINITION 5', mRNA sequence.  
ACCESSION BM472108  
VERSION BM472108.1 GI:18521150  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 1050)  
NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: c9apbs-r@mail.nih.gov  
Tissue Procurement: ATCC/DC/DTP  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLM12233 row: m column: 14  
High quality sequence stop: 658.  
Location/Qualifiers  
1. 1050  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5539381"  
/clone\_lib="NIH\_MGC\_72"



Oy	115	GIUUSerLySgInGlnSerIleGlnGluValThrTPVAlLeuLysAlaPheSer	134
Db	103	GAATTCCTCAACCAACAAGCATTCAGAAGGTACTGGTACTGTTTAAGCGATTCAAGT	162
Oy	135	PheILeArgGLIALGlHisLysSerSerGIuaSnLeuHISProAspAsnValIleLys	154
Db	163	TTTTAAAGGAAAACAGACCCTTAAGAGTTTCAGAAAAATTTGCACCTGACTAATGATAGAA	222
Oy	155	LysLysAsnProPheSerGIUGlyLysPheLysLeuAlaIleGluIleCysIleCysAsn	174
Db	223	AAGAAAACCCATTTTCTGGGAGAAATTCATAACCCAGCTGCCAGAAATTTGCATAAGTAGT	282
Oy	175	GluGluLeuAsnValAsnProGlnAspAsnIGluAsnIleSerTrpThrCysGlnArg	194
Db	283	AAGAGCCTTAATCTTAATCCCAAGACACAGTGGAATAATGTCTCCAGGCTATGTCACAGAGA	342
Oy	195	SerSerGlnGlnSerIleLysSerLeuAlaTrpArgPro---ArgArgLysTrpPheCys	213
Db	343	CCTTCACGGAGCGCCACCCATCACAGGCTCGGAGGCCAGAGAGGAAATGGTTTCAT	402
Oy	214	GlyThrGlyProGlySerLeuCysCysValGlnProArgAspLeuValProCysValPro	233
Db	403	GGCCAGGCCACGAGTCCCTGATTGTGTACAGCCTAGAACACTGGTCCCCTGTGTGCCA	462
Oy	234	ValAsn-----SerIleValAlaSerGlnGly	242
Db	463	GCTTCCCCAGCTGACGTAAAGGGGCCAATGTACAGCTCGGGCTGTGGCTTCAAGAGGT	522
Oy	243	AlaSerProLysProTrpGlnLeuProSerGlyValGlnProValAlaLysLysSer	262
Db	523	GGAAACCCCAAGCCCTGCGAGCTTCACATGTGGTGTGAGCCTCGGGGTCAACAGAAATCA	582
Oy	263	ArgIleGluValTrpGluProProIleArgPheGlnLysIleLysGlnLysProTrpMet	282
Db	583	AGAAATTTGAGTGTGGGAAACCTTTMACTAATTTTCGAAATATGTGGAAATGCTTGATG	642
Oy	283	ProArgGlnLysPheAlaValGlyValGlySerSerTrpArgThrSerAlaArgValVal	302
Db	643	CCAGAGCAAAGTTTCTGTCAGAGAGTGGGAAACCTCATGAGAGAGCCTCTAGGCAAGG	702
Oy	303	GlnLysGlnAsnValGlyTrpGluProProHISArgValProSerGIuAla---ProSe	321
Db	703	TGGAAAGGAAATGTGGCGTTGGAAACCCCACTAATGCTCTAGTGGGCACTCCCTTA	762
Oy	321	rSerArgIlaValAlaArgSerProProSerSerArgLeuGlnLys-GlyArgSerThr	341
Db	763	GTCGAGCTTGTCAAAAAGAGGCGCACATGTCCTCCAGACCCCANNAATGTATGATCCAC	822
Oy	341	sperLeuGlnHis-ValPro-GluLysSerThrAspThrGlnCysGlnPro	357
Db	823	ACAGCTTGGACCGTGGCACCTGGGAAAACCGACGACCTCATGCGCAGCC	874
RESULT 9	B1488505	870 bp	mRNA linear EST 28-AUG-2001
LOCUS	603021014P1 NIH_MGC_114	Homo sapiens cDNA clone IMAGE:5191803 5'	
DEFINITION	mRNA sequence.		
ACCESSION	B1488505		
KEYWORDS	B1488505.1 GI:15327733		
SOURCE	EST.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
JOURNAL	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
COMMENT	NIH-MGC http://mgc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strusberg, Ph.D. Email: cgabbs-remail.nih.gov Tissue Procurement: Life Technologies, Inc. cDNA Library Preparation:Life Technologies, Inc.		

QY	149	ProbspanVal11	lelylsylslysaAsnPropheserGluGlylsPheylsleuAla168
QY	149	ProbspanVal11	lelylsylslysaAsnPropheserGluGlylsPheylsleuAla168
Db	5	CGTGATGATGAGAGAAAGAAACCCTTTT	TGAGAGATAATTCACGACGATGCA64
QY	169	GlutllecylslecyAsnGluGluLeuAsnValasnProGlnAspAsnGlyGluAsnIle188	
Db	65	GAAATTTGCATATGATGACCAAGACGATTAATGTTAATCTCCAGACCATGAGGAAATGTC124	
QY	189	SerTrrphtyGlnArgSerSerGlnGlnSerIleLysSerleuAla1TrpArgPro--A208	
Db	125	TCGGGCCATGTCACAGACCTTCAATGCGACGCCCTCCCATCACA--GGCGAGAGGCCACAGA183	
QY	208	rgArglysrtrPhecyGlyThrGlyProGlySerleucyscyValGlnProArgAspRL228	
Db	184	GGAAGAAAGAGTTTGTGTGGCCAGGCCACAGCATCCCGCTTTGTGACGCTAGGGACT243	
QY	228	euValProCysValProVal-----AsnSerA237	
Db	244	TGCTGCGCTGTCTCCACAGTCGCTCCAGCTGTGACTGAAGGGCCAACTAGAGCTCGGG303	
QY	237	lAlaValAserGlnGlyAlaSerProLysProTrrpGlnleuProSerGlyValGluProV257	
Db	304	CTGTGCTTCACAGAGGTGAGACCTGTAGCTTTGGACGCTTCCACAGTGGTGTGAGCCTA363	
QY	257	AlGlyAlaLysLysSerArgIleGluVal1TrpGluProProLleAsyPheGlnLysIleT277	
Db	364	CAGGTGACGACGAAGTCAAGAACTGAGGTTTGGAAACCTCTGCTCAATTTTCAGAAAGATGT423	
QY	277	yrGlyAsnProTrrpMetProArgGlnLysPheAlaValAlGlyValAlGlySerSerTrrpArgT297	
Db	424	TTGGAAATGCTGCTGATGCCACGACCAAAAGTTTGCCTGACAGAGTGGGCGCTTATTGGAGAA483	
QY	297	hrSerAlaArgValAlGlnLysGlyAsnValAlGlyTrrpGluProProLhisArgValProS317	
Db	484	CCCTGCTGAGAGACGACGAAGAAAGAAATGTGGATTCGAGACCCCGACAGAGATGCATA543	
QY	317	erglyAlaProSerSerArgAlaValArgArgSerProProSerSerArgLeuGlnLysG6337	

```

Db      544 CTGGGGCACTGCTAGTGAAGTGGAGAGACACACCGCTCTCCAGACCCAGAATG 603
      :|||||  |||  |||||||  |||||||||||  |||  |
Qy      337 LYAqserThrAspSerleuGlnHisValProGluLysSerThrAspThrGlnCysGlnP 357
      |||||||  :|||||  |||||||  |||  |||||||  |||
Db      604 GTAqATCCAGCAACGACTTGGACACCTGCTCGAAGAAATGCACAGTACTCATATCCAGC 663
      |||||||  :|||||  |||||||  |||  |||||||  |||
Qy      357 rovalLysAlaAlaGlyMetGluSerValProTyrLysThrValAlaGluLeuThrL 377
      |||||||  |||||||  |||||||  |||  |||||||  |||
Db      664 CCATGAGGCGACCGAGGAGGAGGCTGTACCTCGCAAGCCACAGGAGGAGGCTGTCCA 723
      |||||||  |||||||  |||||||  |||  |||||||  |||
Qy      377 ysrThValGlyLeuLeuHisCysHisAsp-LeuAspValArgHisGlyValLys 396
      |||||||  |||||||  |||||||  |||  |||||||  |||
Db      724 AGACCATGGAAACCA-CTCTGATCAGCGCTGATCGATCGAGACCTGAGTCAAA 782
      |||||||  |||||||  |||||||  |||  |||||||  |||
Qy      397 ArgAspHisPheGlyAlaLeu-ArgPheAspCysProThrGlyPheArgThrTyMetC 416
      |||||||  |||||||  |||||||  |||  |||||||  |||
Db      783 GGAGATCATTTTGGAGCTTAAACATGTGACCTGCTGCTGATTTTCAGACTTGCATTGG 842
      |||||||  |||||||  |||||||  |||  |||||||  |||
Qy      416 Y 416
      |
Db      843 G 843

RESULT 10
BO723415          941 bp  mRNA  linear  EST 16-JUL-2002
LOCUS             BO723415
DEFINITION       AGENCOURT 8305155 Lupski-sympathetic-trunk Homo sapiens CDNA clone
IMAGE:6193072 5', mRNA sequence.
ACCESSION        BO723415
VERSION           BO723415.1 GI:21862312
KEYWORDS          EST.
SOURCE            human.
ORGANISM          Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE        1 (bases 1 to 941)
AUTHORS          NIH-MGC http://mgc.nci.nih.gov/.
TITLE            National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL          Unpublished (1999)
COMMENT          Contact: Robert Strausberg, Ph.D.
                  Email: cgapbs-remail.nih.gov
                  Tissue Procurement: Dr. James R. Lupski
                  CDNA Library Preparation: Life Technologies, Inc.
                  DNA Sequencing by: Agencourt Bioscience Corporation
                  Clone distribution: MGC clone distribution information can be
                  found through the I.M.A.G.E. Consortium/MLN at:
                  http://image.llnl.gov
                  Plate: LHAM13596 row: b column: 17
                  High quality sequence stop: 606.
FEATURES
  source
    1. 941
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:6193072"
      /clone_lib="Lupski-sympathetic-trunk"
      /sex="male"
      /tissue_type="sympathetic trunk"
      /dev_stage="adult, 16 yr"
      /lab_host="DH10B"
      /note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
      NotI; Site_2: SalI; CDNA made by oligo-dT priming.
      Directionally cloned using the following adaptors:
      5'-TCGACCCAGCGGTCG-3' and
      5'-GACTAGTTCTAGATCGGAGCGCGCCCT(15)-3'. Size selected >
      1 kb for average insert length 1.9 kb. This is a primary
      library, non-amplified. Library constructed by Life
      Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
      College of Medicine); available through Life
      Technologies."
BASE COUNT      257 a      267 g      209 t
ORIGIN

```

```

Alignment Scores:
Pred. No.:      1,09e-61      Length:      941
Score:          793.00      Matches:      177
Percent Similarity: 68.51%      Conservative: 21
Best Local Similarity: 61.25%      Mismatches: 73
Query Match:     32.82%      Indels:      18
DB:              14      Gaps:      2

US-09-997-610-2_COPY_18_459 (1-442) x BO723415 (1-941)

Qy      114 GluGluIleSerLysGlnInserIleGlnGluValThrTrpValLeuLeuLysAlaPhe 133
      |||||||  |||||||  |||||||  |||  |||||||  |||
Db      51 GAAGAAATTTCTAAGCAGCAAGCATTTCAAGAGTCTCTTGGGTGCTGTTAAAGACATTC 110
      |||||||  |||||||  |||||||  |||  |||||||  |||
Qy      134 SerPheIleArgGluAlaGluHisLysSerSerGluAsnLeuHisProAspAsnValIle 153
      |||||||  |||||||  |||||||  |||  |||||||  |||
Db      111 AGTTTTAAAGGGAACACAGACATAAAGTTTGGAAATTTTCACGCTGCAATGCAATA 170
      |||||||  |||||||  |||||||  |||  |||||||  |||
Qy      154 LysLysLysAsnProPheSerGlu-GlyLysPheLysLeuAlaGluIleCysIleCys 173
      ::|||  |||  |||  |||  |||  |||
Db      171 GAAAGAAAATCCCATTTTCTGAGGAGAAAGTCAAGCTGCTACAGAAATTTACTAATAG 230
      |||||||  |||||||  |||||||  |||  |||||||  |||
Qy      173 sAsnGluLeuAsnValAsnProGlnAspAsn-GlyGluAsnIleSerTrpThrCysG 193
      |||||||  |||||||  |||||||  |||  |||||||  |||
Db      231 TAACAAGAAAGCCAAATGTTAATCCCAAGACATAGGGGTAAATGCTCCAGAGCATATC 290
      |||||||  |||||||  |||||||  |||  |||||||  |||
Qy      193 InArGserSerGlnGlnInserIleLysSerLeuAlaThrParProArgArg---LysTrp 212
      |||||||  |||||||  |||||||  |||  |||||||  |||
Db      291 AGAGGTCTTTCACGCGACGCCCTCCCTTTCACAGGCTGGGGGCTTAGAGAGAAAGTGGT 350
      |||||||  |||||||  |||||||  |||  |||||||  |||
Qy      212 heCysGlyThrGlyProGlySerLeuCysCysValGlnProArgAspLeuValProCysV 232
      ||  |||  |||  |||  |||  |||
Db      351 TTCGTGGGCTCAGACAGGGGTCCCATGCTGTGTGACGCTTAGGGACTGTGGCTTGA 410
      |||||||  |||||||  |||||||  |||  |||||||  |||
Qy      232 aProVal-----AsnSerAlaValAlaSerG 241
      |||  |||  |||  |||  |||  |||
Db      411 TCACAGCTGCTCAGCTGTGGCTGAAGGGGCTATCACAGACTCAGCGCTGTGGCTTCA 470
      |||||||  |||||||  |||||||  |||  |||||||  |||
Qy      241 IuGlyAlaSerProLysProTrpGlnLeuProSerGlyValGluProValGlyAlaLysL 261
      |||||||  |||||||  |||||||  |||  |||||||  |||
Db      471 AGGGTGCAAGTCCCAAGGCTTGGCAGCTTCCACGTGGTATTGAGCTGCAGATACATGCA 530
      |||||||  |||||||  |||||||  |||  |||||||  |||
Qy      261 ySerArgIleGluValTrpGluProIleArgPheGlnLysIleTyrGlyAsnProT 281
      |||||||  |||||||  |||||||  |||  |||||||  |||
Db      531 AGTCAAGAAATTTGGGTGGAACTCTGCTAGATTTTTGGAGAGTACGGAATAACT 590
      |||||||  |||||||  |||||||  |||  |||||||  |||
Qy      281 rMetProArgGlnLysPheAlaValGlyLysSerSerTrpArgTrpSerAlaArgV 301
      |||||||  |||||||  |||||||  |||  |||||||  |||
Db      591 GGATGCCCGAGAGAAAGTTGTTGCGAGGACAGGGGCCCTCATGAGAACTTCTCGTAGAG 650
      |||||||  |||||||  |||||||  |||  |||||||  |||
Qy      301 aLValGlnLysGlnAsnValGlyTrpGluProProHisArgValProSerGlyAla-Pro 320
      |||||||  |||||||  |||||||  |||  |||||||  |||
Db      651 TAGTGGGAGGAGGAATGATGAGGTGACAGCCCCACACAGAGTCCCTACTGAGACACTGC 710
      |||||||  |||||||  |||||||  |||  |||||||  |||
Qy      321 SerSerArgAlaValAlaArgArgSerProProSerSerArgLeuGlnLys-GlyArgSerTh 340
      |||  |||  |||  |||  |||  |||
Db      711 CCTAGTGGAGCTGTGAGAAAGCCATCATCTTTCACACCCAGGAAAGTGTAGATCCAC 770
      |||  |||  |||  |||  |||  |||
Qy      340 rAspSer-LeuGlnHisValProGluLysSerThrAspThrGlnCysGlnProValLysA 360
      |||||||  |||  |||  |||  |||  |||
Db      771 TACACGCTTGCACGTGTGCTCGAAGAACACACACACTCAAGCCACGACCATTAAG 830
      |||  |||  |||  |||  |||  |||
Qy      360 Ia-AlaGlyMetGluSerValProTyrLysThrValAlaGluLeuThrLysThrVal 379
      ||  ||  ||  ||  ||  ||
Db      831 CACCCAGGAGGAGGAGCTGTACCTTCGCAAGACACAGGGGGTGGGACACCCCAAGACTAT 890
      ||  ||  ||  ||  ||  ||
Qy      380 GlyIleTyrLeuLeu 384
      ||  ||  ||  ||
Db      891 GGGACCAACCTCTT 905
      ||  ||  ||  ||

RESULT 11
BM554723          1076 bp  mRNA  linear  EST 20-FEB-2002
LOCUS

```

DEFINITION AGENCOURT 6578521 NIH\_MGC\_41 Homo sapiens cDNA clone IMAGE:5468842  
5', mRNA sequence.  
ACCESSION BM554723  
VERSION BM554723.1 GI:18794564  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1076)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
AUTHORS Contact: Robert Strausberg, Ph.D.  
JOURNAL Email: cga@bbs-remail.nih.gov  
COMMENT Tissue Procurement: DCTD/DP  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L101 row: b column: 11  
High quality sequence stop: 665.  
location/Qualifiers  
1. 1076  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="NIH\_MGC\_41"  
/tissue\_type="amelanotic melanoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pORF7; Site:1: XhoI; Site:2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGCAG(6). Library constructed by Ling Hong in the  
Laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC library."  
BASE COUNT 274 a 251 c 295 g 256 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 6.44e-58 Length: 1076  
Score: 752.50 Matches: 154  
Percent Similarity: 68.75% Conservative: 11  
Best Local Similarity: 64.17% Mismatches: 62  
Query Match: 31.15% Indels: 13  
Gaps: 1  
US-09-997-610-2\_COPY\_18\_459 (1-442) x BM554723 (1-1076)  
QY 216 GYTPGGLYSerLeuCySValGlnProArGAspleuValProCySValProValasn 235  
|||||  
Db 5 GGCCCAAGGCTCCCTGTCTGTGTACAGGCTAGGAGCTTGGTGTCCAGCCACT 64  
QY 236 SerAla-----ValAlaSerGluGlyAlaSer 244  
|||||  
Db 65 CCAGCCATGGGTGAAGGACCAACATAGAGCTAGAGCTGTGCTTCAGAGATCCAAAG 124  
QY 245 ProLysProTyrPglInleuProSerGlyValGlnProValGlyAlaLysSerArgile 264  
|||||  
Db 125 CCCAAGTCTGGCAGCTTCCACATGAGTGTGAGCCTCAGATGCACATAGTCAAGATT 184  
QY 265 GluValTTPGlnProProLleArgPheGlnLysLleTyrGlyAsnProTyrMetProArg 284  
|||||  
Db 185 GGAGTTTGGAACCTCCCTAGATTTCAGAGATGTATGGAATGCTGGATGTCAGG 244  
QY 285 GlnLysPheAlaValGlnValGlySerSerTyrPArGthrSerAlaArgValGlnLys 304  
|||||  
Db 245 AAGAAAGTGGCTACAGGGGACAGGCACTCATGAGAACTCTGCTAGACAGTGCAGAG 304

QY 305 GlyAsnValGlyTyrPgluProProHisArgValProSerGlyAlaProSerSerArgala 324  
|||||  
Db 305 GGAATATGGTGGGTACAGACCCCAACAGAGCTCTCATTTGGGCACTAGGCTATGGAGGT 364  
QY 325 ValArgArgSerProProSerSerArgLeuGlnLysGlyArgSerThrAspSerLeuGln 344  
|||||  
Db 365 ATGAGAAAGGGCCATCTACTCGAGACCCCAAGCATGTGATGTCACCCACACTTTCGAC 424  
QY 345 HisValProGluLysSerThrAspThrGlnCysGlnProValLysAlaAlaGlyMetGlu 364  
|||||  
Db 425 TGTGCACCTGGAAAGCCACACACACTCAATGCCACCATGAAAGCGGTGAGAGGAG 484  
QY 365 SerValProTyrLysThrValValAlaGluLeuThrLysThrValGlyLysLeu 384  
|||||  
Db 485 GCTGTACCTTACAAAGCCACAGGACAGAGTCGCCCAAGCATGTGAAACCCACTTTCG 544  
QY 385 HisCysHisAspLeuAspValArgHisGlyValLysArgAspHisPheGlyAlaLeuArg 404  
|||||  
Db 545 CAAACACATGACTCGGATGTGAGACATGAGTCAAAGAGATTAATTTTGGAGCTTTAAG 604  
QY 405 PheAspCysProThrGlyPheArgThrTyrMetGlyProVal-ProLeuCySPheGly 424  
|||||  
Db 605 TTTCAGTCCCACTGAGATTTTCGACTGCTTCGCGCTGAGCCCTTTGTTGGCCA 664  
QY 424 nPhePhePro-PheGlyThrAlaValPheThrGlnCysLeuTyrLeuHisCysMet 442  
|||||  
Db 665 ATTTCACATGTGGGAATGGCTGTATTACTCAATGCTCAATCCCAATTGATC 720  
RESULT 12  
BQ422247/c 1080 bp mRNA linear EST 23-MAY-2002  
LOCUS BQ422247  
DEFINITION AGENCOURT\_7766790 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:6015590  
5', mRNA sequence.  
ACCESSION BQ422247  
VERSION BQ422247.1 GI:21117562  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1080)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
AUTHORS Contact: Robert Strausberg, Ph.D.  
JOURNAL Email: cga@bbs-remail.nih.gov  
COMMENT Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L101 row: o column: 15  
High quality sequence stop: 702.  
location/Qualifiers  
1. 1080  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="NIH\_MGC\_92"  
/tissue\_type="embryonal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: testis; Vector: pCMV-SPORT6; Site:1: NotI;  
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 2.5 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC library."  
BASE COUNT 239 a 304 c 221 g 316 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 2.53e-56 Length: 1080





TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov

Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:  
http://image.lnl.gov  
Plate: LLM1214 row: a column: 13  
High quality sequence stop: 599.

## FEATURES

source

Location/Qualifiers  
1. 998  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5492244"  
/clone\_lib="NIH\_MGC\_67"  
/tissue\_type="retinoblastoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: PCMV-SpOFT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo  
Average insert size 1.75 kb. Library constructed by Life  
Technologies."

BASE COUNT 252 a 241 c 284 g 219 t 2 others  
ORIGIN

## Alignment Scores:

Pred. No.: 5,19e-56 Length: 998  
Score: 731.00 Matches: 183  
Percent Similarity: 64.65% Conservative: 20  
Best Local Similarity: 58.28% Mismatches: 83  
Query Match: 30.26% Indels: 5  
DB: 14 Gaps: 5

US-09-997-610-2\_COPY\_18\_459 (1-442) x BM903785 (1-998)

QY 142 LysSerSerGluAsnLeuHisProAspAsnValIleLysLysLysAsn--ProPheSerG 161  
Db 2 AAAAGTTGGAAATTTTCACGCTGACAAATGATAGAAAGAAATTCCTATTTTCG 61  
QY 161 LglLysPheLysLeuAlaIleGluIleCysIleCysAsnGluLysLeuAsnValAsp 181  
Db 62 AGGAGAAATTCAGCTGGCTGCAGAAATTTGCATGATACAG-GAGCCAAATGCTAAAT 120  
QY 181 roGlnAspAsnGlyGluAsnIleSerTrpTrpCysGlnArgSerSerGlnGlnIleL 201  
Db 121 CCCAAGACAAATGGGGAATATCTCTCAGGCGCATGTCCAGAGCTTTATGGCAACCCCTCC 180  
QY 201 ySerLeuAlaTrpArg-ProArgArgLysTrpPheCysGlyThGlyProGlySerLeu 220  
Db 181 CATCACAGCTCCAGATATCAGAAATAATGTTTGTGGCCAGGCCGGGCTCTCA 240  
QY 221 CysCysValGlnProArgAspLeuValProCysValProVal----- 234  
Db 241 TGCTGTGTGACGCTTAGGACTTGGTCCCTGCATCCAGCCACCTCCCAACCATGACTGA 300  
QY 235 -----AsnSer-AlaValAlaSerGluLysAlaSerProLysProTrpG 249  
Db 301 CAGAGGCAAGTAGAGCTTGGGCTGTAGCTTGGGGAGTGCAGACCCCAACGCTTGACA 360  
QY 249 nLeuProSerGlyValGluProValGlyAlaLysSerArgIleGluValTrpGluPr 269  
Db 361 GCTTCCATGTGTGTGAGACTGCGAGTGACACAGAACTGAGAACTGGGGTTTGAAAC 420  
QY 269 oProLleArgPheGlnLysIleTrpGlyAsnProTrpMetProArgGlnLysPheAlaVa 289  
Db 421 TTCCGCTAGATTAAAGAGAGTGTGCGAAATGCTTGATGCGCAGGCAAGATTGCTGC 480  
QY 289 lGlyValLysSerTrpArgTrpSerAlaArgValValGlnLysGlyAsnValGlyTr 309  
Db 481 AGGGGCAAGGCGCTCATGTGAGATCTCTGCGCAGGCGCATGCGAAGGGAATGTGGGCTC 540

QY 309 pGluProHisArgValProSerGlyAlaProSerSerArgAlaValArgSerPr 329  
Db 541 AGAGACCCACACACAGTCCCTACTGGGCGACCCACTAGTGAAGTGAAGAGGTCC 600  
QY 329 oProSerSerArgLeuGlnLysGlyArgSerTrpAspSerLeuGlnHisValProGluL 349  
Db 601 TCCAGACCC-----CAGAAATGTAGATCCACCGACCTTGCAACCGTGTACCTGGAAA 653  
QY 349 sSerThrAspTrpGlnCysGlnProValLysAlaIleGlyMetGluSerValProTrp-L 369  
Db 654 AGTTCAGACACTGCATGCTGCACGCCCATGMAAGCAGCTGAGAGGAGGCTGACCCCTGCA 713  
QY 369 yThrValValAlaIleGluLeuThrLysThr-ValGlyIleTrp-LeuLeuHisCysHisAs 388  
Db 714 AGGTACAGGGGCGAGAGTGGCCCAAGACCATGGGAACCCACCTTNCATCTGCAATGA 773  
QY 388 pLeuAspValArgHisGly-----ValLysArgAspHis-PheGlyAlaLeuArg-Phe 405  
Db 774 CTTGGATGTGAAATGTGTGAGTGCAGTGCAGAGGAAAGATCTTTTGTGAGCTTTTAGATTT 833  
QY 406 AspCysProThr---GlyPheArgThrTyMetGlyPro-----ValProLeuCysPhe 422  
Db 834 GACTGGCCCCCTGGGATTTGTGACTCTCATGGGCGCTTGGAAACCTCTTGGTTTGG 893  
QY 423 GlyGlnPhe-PheProPheGly 429  
Db 894 ACCCAATTTATTCCTCAATTTGGG 915

Search completed: February 19, 2003, 19:35:44  
Job time : 3455.83 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 19, 2003, 14:57:18 : Search time 6158.57 Seconds  
(without alignments)  
2088.706 Million cell updates/sec

Title: US-09-997-610-2\_COPY\_18\_459  
Perfect score: 2416  
Sequence: 1 AGPRAHRRPPEVGPARGA.....GQFFPGTAVFTQCILYHGM 442

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+pn.model -DEV=xlh  
-O=/cgn2.1/USPTO/spool/US09997610/runat\_10022003\_160823\_23819/app\_query.fasta\_1.1635  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blonsun62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09997610@cgn2.1\_13965@runat\_10022003\_160823\_23819 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMECUT=120  
-WARN\_TIMECUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*  
1: gb\_da.\*  
2: gb\_hlg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

29: em\_vl.\*  
30: em\_hlg\_hum.\*  
31: em\_hlg\_inv.\*  
32: em\_hlg\_other.\*  
33: em\_hlg\_mus.\*  
34: em\_hlg\_pin.\*  
35: em\_hlg\_rod.\*  
36: em\_hlg\_mam.\*  
37: em\_hlg\_vrt.\*  
38: em\_sy.\*  
39: em\_hlgo\_hum.\*  
40: em\_hlgo\_mus.\*  
41: em\_hlgo\_other.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2327	96.3	145880	9	HS302D9	282198 Human DNA s
2	1218.5	50.4	161575	2	AC025233	AC025233 Homo sapi
3	1218.5	50.4	171625	2	AC087500	AC087500 Homo sapi
4	1217.5	50.4	123098	9	AC008799	AC008799 Homo sapi
5	1217.5	50.4	174874	9	AC123595	AC123595 Homo sapi
6	1217.5	50.4	189768	2	AC044889	AC044889 Homo sapi
7	1209	50.0	91323	9	AC106901	AC106901 Homo sapi
8	1207	50.0	112515	9	AL139216	AL139216 Human DNA
9	1205	49.9	137785	9	HS345P10	282201 Human DNA s
10	1200	49.7	168080	2	AC009047	AC009047 Homo sapi
11	1200	49.7	199579	9	AC093536	AC093536 Homo sapi
12	1199.5	49.6	155764	9	AC011238	AC011238 Homo sapi
13	1199.5	49.6	187898	2	AC016715	AC016715 Homo sapi
14	1198	49.6	99084	2	AC026420	AC026420 Homo sapi
15	1198	49.6	133790	9	AC010280	AC010280 Homo sapi
16	1198	49.6	170723	9	AC013470	AC013470 Homo sapi
17	1198	49.6	174293	9	AC008816	AC008816 Homo sapi
18	1198	49.6	174293	9	AC010242	AC010242 Homo sapi
19	1198	49.6	203043	9	AC017093	AC017093 Homo sapi
20	1195.5	49.5	186660	2	AC026107	AC026107 Homo sapi
21	1195	49.5	40714	9	HS0212C1	269722 Human DNA s
22	1194.5	49.4	124518	9	AL138962	AL138962 Human DNA
23	1194	49.4	166757	9	AC003046	AC003046 Homo sapi
24	1194	49.4	185171	2	AC023098	AC023098 Homo sapi
25	1193.5	49.4	174380	2	AC016075	AC016075 Homo sapi
26	1193.5	49.4	175559	9	AC112498	AC112498 Homo sapi
27	1193	49.4	68935	9	AC008150	AC008150 Homo sapi
28	1193	49.4	77405	2	AL353634	AL353634 Homo sapi
29	1192	49.4	176784	2	AC016632	AC016632 Homo sapi
30	1192	49.3	182126	2	AC016639	AC016639 Homo sapi
31	1190.5	49.3	154090	2	AC025577	AC025577 Homo sapi
32	1190	49.3	149080	2	AC087699	AC087699 Homo sapi
33	1187	49.1	168502	9	AC091005	AC091005 Homo sapi
34	1186.5	49.1	172919	9	AL807742	AL807742 Human DNA
35	1185.5	49.1	123631	9	HS22F01	AL109967 Homo sapi
36	1184.5	49.0	94730	9	AP000230	AP000230 Homo sapi
37	1184.5	49.0	100000	9	AP000144	AP000144 Homo sapi
38	1184.5	49.0	100634	9	AP001594	AP001594 Homo sapi
39	1184.5	49.0	182778	2	AC022312	AC022312 Homo sapi
40	1184.5	49.0	340000	9	AP001635	AP001635 Homo sapi
41	1184	49.0	173480	9	CNS00WGT	AL079343 Human chr
42	1183	48.9	159875	9	AC024023	AC024023 Homo sapi
43	1182.5	48.9	153223	2	AC074339	AC074339 Homo sapi
44	1182	48.9	110820	9	AL596285	AL596285 Human DNA
45	1181	48.9	182972	2	AC023550	AC023550 Homo sapi

RESULT 1

## ALIGNMENTS

HS302D9 145880 bp DNA linear PRI 12-DEC-1999  
LOCUS HS302D9  
DEFINITION Human DNA sequence from clone RP1-302D9 on chromosome 22 Contains  
GSSs, complete sequence.  
ACCESSION 282198  
VERSION 282198.2 GI:6572207  
KEYWORDS HMG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 145880)  
AUTHORS Bridgeman, A.  
TITLE Direct Submission  
JOURNAL Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
Requests: clonerequest@sanger.ac.uk  
On Dec 13, 1999 this sequence version replaced gi:3164067.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
Chromosome 22, constructed by the Sanger Centre Chromosome 22  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr22  
RP1-302D9 is from the library RPIC1-1 constructed at the Roswell  
Park Cancer Institute by the group of Pieter de Jong. For further  
details see http://bacpac.med.buffalo.edu/  
VECTOR: pCYPAC2  
This sequence is the entire insert of clone RP1-302D9 The true left  
end of clone CTA-282F2 is at 69682 in this sequence. The true right  
end of clone CTA-415G2 is at 55167 in this sequence.  
FEATURES  
source  
1..145880  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="22"  
/clone="RP1-302D9"  
/clone\_id="RPIC1-1"  
188..245  
/note="MER3 repeat: matches 144..209 of consensus"  
246..571  
/note="Alusx repeat: matches 1..312 of consensus"  
572..759  
/note="MER3 repeat: matches 1..144 of consensus"  
783..933  
/note="MER5A repeat: matches 26..187 of consensus"  
1033..1336  
/note="AlusP repeat: matches 1..299 of consensus"  
1450..1583  
/note="MIR repeat: matches 24..160 of consensus"  
1687..1752  
/note="L2 repeat: matches 2593..2661 of consensus"  
2350..2660  
/note="AlusC repeat: matches 3..309 of consensus"  
2684..2981  
/note="AlusQ repeat: matches 2..300 of consensus"  
3323..3343  
/note="MIR1E repeat: matches 116..136 of consensus"

repeat\_region 3344..3652  
/note="AluY repeat: matches 1..309 of consensus"  
3553..3928  
/note="MIR1E repeat: matches 136..359 of consensus"  
3929..4278  
/note="MER1B repeat: matches 3..364 of consensus"  
4279..4485  
/note="MIR1E repeat: matches 359..568 of consensus"  
5073..5176  
/note="52 copies 2 mer ct 78 conserved"  
5181..5491  
/note="AluDb repeat: matches 1..311 of consensus"  
6369..6485  
/note="L2 repeat: matches 2579..2705 of consensus"  
6647..6685  
/note="MADE1 repeat: matches 1..23 of consensus"  
6686..6987  
/note="AlusX repeat: matches 1..302 of consensus"  
6988..7036  
/note="MADE1 repeat: matches 23..77 of consensus"  
7482..7754  
/note="AluDb repeat: matches 9..290 of consensus"  
7775..8060  
/note="AluDo repeat: matches 1..295 of consensus"  
8414..8551  
/note="L2 repeat: matches 2553..2706 of consensus"  
8914..9030  
/note="MIR repeat: matches 147..262 of consensus"  
9110..9280  
/note="MIR repeat: matches 91..262 of consensus"  
9283..9412  
/note="MIR repeat: matches 15..144 of consensus"  
9521..9679  
/note="FAM repeat: matches 3..161 of consensus"  
9820..10225  
/note="MSTR repeat: matches 2..425 of consensus"  
complement(10179..10678)  
/note="match: GSS: Em:B56592"  
complement(10204..10728)  
/note="match: GSS: Em:AQ701486"  
complement(10249..10706)  
/note="match: GSS: Em:AQ225495"  
10312..10383  
/note="MIR repeat: matches 79..150 of consensus"  
10718..11310  
/note="match: GSS: Em:B14024"  
10784..11201  
/note="match: GSS: Em:B43656"  
11838..11946  
/note="MIR repeat: matches 20..137 of consensus"  
12174..12445  
/note="L2 repeat: matches 1988..2275 of consensus"  
12444..12642  
/note="MIR repeat: matches 63..241 of consensus"  
13017..13369  
/note="match: STS: Em:G49301"  
13331..13397  
/note="MIR repeat: matches 174..244 of consensus"  
13398..13698  
/note="AlusP repeat: matches 1..302 of consensus"  
13699..13810  
/note="MIR repeat: matches 76..174 of consensus"  
13806..13919  
/note="MIR repeat: matches 77..189 of consensus"  
13945..14060  
/note="MIR repeat: matches 24..142 of consensus"  
14061..14367  
/note="AluY repeat: matches 1..301 of consensus"  
14368..14452  
/note="MIR repeat: matches 141..225 of consensus"  
14589..14679  
/note="MIR repeat: matches 173..262 of consensus"  
14597..15201  
misc\_feature

QY	56	PhelysProIleIlePheThrGlyValLeuTyrrAsnAlaGlnArgAspLeuLysGluAla	75
Db	37382	TTCAAGCCCATCATCTTTCACAGGGGCTCGTACAAATGCCACAGAGGATTTAAAGAGGCC	37441
QY	76	MetGlyValAlaPheAlaCysArgValProGlyAsnTyrrTyrSerSerPheAspValGluLeu	95
Db	37442	ATGGAGAGTCCTTTCCTTGCACGAGGCTCGGGAATTAATCTCCAGCTTTCATGTTCAGCGTG	37501
QY	96	HisHisCysLysValAsnIleTrpLeuMetArgLysGlnIleLeuAlaAsnLysGluGln	115
Db	37502	CATCATTCGACAGAGTGAATATTTGGCTAATGAGAGCAAAATTTGGCTAATTAAGCAAGAA	37561
QY	116	IleSerLysGlnGlnSerIleGlnGluValIrrTrpValLeuLeuLysAlaPheSerPhe	135
Db	37562	ATTTCCTAACAGCAACAAAGCATTCACAGAGTGAGCTGGGTGCTGTAAAGCATTCAGATTTC	37621
QY	136	IleArgGluAlaGluHisLysSerSerGluAsnLeuHisProAspAsnValIleLysLys	155
Db	37622	ATAAGGAGAGCGAGAGCATTAAGAGTTCACAAAAATTTGCACCCCTGCACATGTGATAAAAAG	37681
QY	156	LysAsnProPheSerGlnGlnLysPheLysLysPheLysLeuAlaAlaGlnIleCysIleCysAsnGln	175
Db	37682	AAAAACCCATTTTCTGAGGGGAAATTCAAAGCTGGCTGGCGCAAAATTTGCATATGTAATGAG	37741
QY	176	GluLeuAsnValAsnProGlnAspAsnGlyGluAsnIleSerTrpThrCysGlnArgSer	195
Db	37742	GAGCTGAATGTTAATCCTCAACAGCATAGGGGAAATATCTCCGTGACATGTACAGAGTCT	37801
QY	196	SerGlnGlnSerIleLysSerLeuAlaTrpArgProArgValArgLysTrpPheCysGlyThr	215
Db	37802	TCACAGCACTCATCAAAATCACTGGCTTGAGAGCCTAGAGAGAAATGGTTTGTGTGGACA	37861
QY	216	GlyProGlySerLeuCysCysValGlnProArgAspLeuValProCysValProValAsn	235
Db	37862	GGCCAGAGGCTCCCTGTCTGTGTGTCAGCCCTAAGAGACTTGCTCCCTGTCTCCAGTTAAT	37921
QY	236	SerAlaValAlaSerGlnGluLysAspProLysProTrpGlnLeuProSerGlyValGln	255
Db	37922	TCAGCTGTGCTTCACAGAGGTCGCAAGCCCAAGCCCTTGCCACTTCGCAAGTGTTGTAG	37981
QY	256	ProValGlyAlaLysLysSerArgIleGluValTrpGluProProIleArgPheGlnLys	275
Db	37982	CCTGTGGGTGCACAAAGAGTCAGAAATTTGAGTTTGGAACTCCCATCAGATTCGGAAG	38041
QY	276	IleTyrrGlyAsnProTrpMetProArgGlnLysPheAlaValGlyValGlySerSerTrp	295
Db	38042	ATATATGGAAACCCCTGGATGGCCAGGCGAAAGTTTGGCTGAGGGGTGGGTCCTCATAG	38101
QY	296	ArgTrnSerAlaArgValValGlnLysGlyAsnValGlyTrpGluProProHisArgVal	315
Db	38102	ACAACCTCTGCAGAGGTACTACAAAAAGGAATGTTGGGTGGGAGCCCCACACAGAGTC	38161
QY	316	ProSerGlyAlaProSerSerArgAlaValArgArgSerProProSerSerArgLeuGln	335
Db	38162	CCCACTGGGGCTCCATCTAGTAGAGCTGTGAAGAAGATCCACCATCTCCAGAGCTCCAG	38221
QY	336	LysGlyArgSerThrAspSerLeuGlnHisValProGluLysSerThrAspPheGlnCys	355
Db	38222	AAGGTAGATCCACGTACAGCTTGGACGTGGCCCTGAAAAATCCACACACACTCAGTCC	38281
QY	356	GlnProValLysAlaAlaGlyMetGluSerValProTyrrLysThrValValAlaGluLeu	375
Db	38282	CAGCTGTGAAGACACAGCAGGAGTGGAGTGTACCTTACCAAAACCGTAGTGGCAGAGCTG	38341
QY	376	ThrLysThrValGlyIleTyrrLeuLeuHisCysHisAspLeuAspValArgHisGlyVal	395
Db	38342	ACCAGAGCCGTTGGATCTACCTTTCATGTCTACAGACCTGGAGCTGAGACATGAGAGTC	38401
QY	396	LysArgAspHisPheGlyAlaLeuArgPheAspCysProThrGlyPheArgThrTyrrMet	415
Db	38402	AAAAAGATCATTTTGGAGCTTTAAGATTTCAGTCCGCCACAGTGAGATTTCGGAGCTATATAG	38461

QY 416 GlyProValProLeuCySPheGlyGlnPhePheProPheGlyThrAlaValPheThrGln 435  
Db 38462 GGCCCGATACCCCTTTGTTTGGCAATTTTCCATTGTGGAATGCCGTAATTACCCAA 38521  
QY 436 CysLeuTyrlleuHisCysMet 442  
Db 38522 TGCCTGTACCTCATTGATG 38542  
RESULT 2  
AC025233/c 161575 bp DNA linear HTG 07-JUL-2000  
LOCUS AC025233  
DEFINITION Homo sapiens chromosome 17 clone RP11-333E1, WORKING DRAFT  
ACCESSION AC025233  
VERSION AC025233.4 GI:8954317  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 161575)  
AUTHORS Waterston,R.H.  
JOURNAL Unpublished  
TITLE The sequence of Homo sapiens clone  
REFERENCE 2 (bases 1 to 161575)  
AUTHORS Waterston,R.H.  
JOURNAL Direct Submission  
TITLE Submitted (07-MAR-2000) Genome Sequencing Center, Washington  
REFERENCE University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
COMMENT On Jul 7, 2000 this sequence version replaced gi:7801489.  
----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
----- Project Information -----  
Center project name: H.NH033E01  
----- Summary Statistics -----  
Sequencing vector: M13; 90%  
Sequencing vector: plasmid; 10%  
Chemistry: Dye-Primer ET; 81% of reads  
Chemistry: Dye-terminator Big Dye; 19% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 144558 bases at least Q40  
Consensus quality: 149743 bases at least Q30  
Consensus quality: 152625 bases at least Q20  
Insert size: 176000; agarose-fp  
Insert size: 158975; sum-of-ctnigs  
Quality coverage: 3.48 in Q20 bases; agarose-fp  
Quality coverage: 3.75 in Q20 bases; sum-of-ctnigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 27 ctnigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the ctnigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1141: contig of 1141 bp in length  
\* 1142 1241: gap of unknown length  
\* 1242 2424: contig of 1183 bp in length  
\* 2425 2524: gap of unknown length  
\* 2525 3961: contig of 1437 bp in length  
\* 3962 4061: gap of unknown length  
\* 4062 5712: contig of 1651 bp in length  
\* 5713 5812: gap of unknown length  
\* 5813 8019: contig of 2207 bp in length  
\* 8020 8119: gap of unknown length  
\* 8120 10208: contig of 2089 bp in length  
\* 10209 10308: gap of unknown length  
\* 10309 11734: contig of 1426 bp in length

11735 11834: gap of unknown length  
\* 11835 13265: contig of 1431 bp in length  
\* 13266 13365: gap of unknown length  
\* 13366 14425: contig of 1060 bp in length  
\* 14426 14525: gap of unknown length  
\* 14526 15609: contig of 1084 bp in length  
\* 15610 15709: gap of unknown length  
\* 15710 17529: contig of 1820 bp in length  
\* 17530 17629: gap of unknown length  
\* 17630 20351: contig of 2722 bp in length  
\* 20352 20451: gap of unknown length  
\* 20452 23753: contig of 3302 bp in length  
\* 23754 23853: gap of unknown length  
\* 23854 26537: contig of 2684 bp in length  
\* 26538 26637: gap of unknown length  
\* 26639 31339: contig of 4702 bp in length  
\* 31340 31439: gap of unknown length  
\* 31440 37041: contig of 5602 bp in length  
\* 37042 37141: gap of unknown length  
\* 37142 45268: contig of 8127 bp in length  
\* 45269 45368: gap of unknown length  
\* 45369 52323: contig of 6955 bp in length  
\* 52324 52423: gap of unknown length  
\* 52424 60239: contig of 7816 bp in length  
\* 60240 60339: gap of unknown length  
\* 60340 68385: contig of 8046 bp in length  
\* 68386 68485: gap of unknown length  
\* 68486 78834: contig of 10349 bp in length  
\* 78835 78934: gap of unknown length  
\* 78935 90422: contig of 11488 bp in length  
\* 90423 90522: gap of unknown length  
\* 90523 100763: contig of 10241 bp in length  
\* 100764 100863: gap of unknown length  
\* 100864 111672: contig of 10809 bp in length  
\* 111673 111772: gap of unknown length  
\* 111773 126855: contig of 15083 bp in length  
\* 126856 126955: gap of unknown length  
\* 126956 142557: contig of 15602 bp in length  
\* 142558 142657: gap of unknown length  
\* 142658 161575: contig of 18918 bp in length.  
FEATURES  
source 1..161575  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/clone="RP11-333E1"  
BASE COUNT 44288 a 35882 c 35114 g 43686 t 2605 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 6.53e-56 Length: 161575  
Score: 1218.50 Matches: 239  
Percent Similarity: 74.17% Conservative: 28  
Best Local Similarity: 66.39% Mismatches: 77  
Query Match: 50.43% Indels: 16  
DB: 2 Gaps: 3  
US-09-997-610-2\_COPY\_18\_459 (1-442) x AC025233 (1-161575)  
QY 97 HisCysLysValAsnIleTrp-----LeuMetAlaGlyGlnIleLeuAlaAsnLys 113  
Db 121671 CACTGCCCTAGAGATTTGTGGCACTTTGAACSTTGAGAGAGATGATTGAGGTATCTGGTG 121612  
QY 114 GlnGluIleSerIysGlnGlnIleGlnGlnIleValThrTrpValLeuLeuAlaPhe 133  
Db 121611 GAAAGAAATTTCTAAGCAGCAAAACATTTCAAGGTGTCACTTGAGCTGTGTAATGAGATTG 121552  
QY 134 SerPheIleArgGlnAlaGlnIleIleIleSerSerGlnAsnLeuHisProAspAsnValIle 153  
Db 121551 AGTTTATGAGGAGAGAGAGCATAAAGTTTCAGAAATTTGCAGACCTGCAGAAATGTGATA 121492  
QY 154 LysLysLysAsnProPheSerGlnGlyLysPheLysLeuAlaAlaGlnIleCysIleCys 173  
:::|||||



```

FEATURES                                Location/Qualifiers
source                                  1..171625
                                         /organism="Homo sapiens"
                                         /db_xref="taxon:9606"
                                         /chromosome="17"
                                         /map="17"
                                         /clone="RP11-333E1"
                                         /clone_id="RP11 Human Male BAC"
BASE COUNT      47416 a 39015 c 37523 g 47508 t 163 others
ORIGIN

Alignment Scores:
Pred. No.:      6.99e-56      Length:      171625
Score:          1218.50      Matches:      239
Percent Similarity: 74.17%      Conservative: 28
Best Local Similarity: 66.39%      Mismatches: 77
Query Match:    50.43%      Indels:      16
DB:             2      Gaps:      3

```

OY	97	HisCysValValAsnIleTrp-----LeuMetArgValGlnIleLeuAlaAsnLys	113
Db	136713	CACGGCCCTAGAGAAATTTTGGAACTTGGAACTTGAACATGAGAGATGATTTTGGAGTATCTGGCTG	136772
OY	114	GIUGLUILeSerLysGlnGlnSerTlEglnGluValThpTrpValIleuLeuLysAlaPhe	133
Db	136773	GAGGAATTTCTTAAGACAGAAACATTCAAGGTGTGACTTGGGTGCTCTTAATGAGATTC	136832
OY	134	SerPheIleArgGluAlaGlnHisLysSerSerGluAsnLeuHisProAspAsnValIle	153
Db	136833	AGTTTATATAGAGGAAGACAGACACATTAAGAAATTCAGAAATTTTCACCCGACAAAGTGATA	136892
OY	154	LysIleLysAsnProPheSerGluGlyLysPheLysLeuAlaGlnIleCysIleCys	173
Db	136893	GAAGAAGAAACCCATTTTCTGAGAGAAATTTCAATCAGGCTGCAGAAATTTGGATAGT	136952
OY	174	AsnGluGluLeuAsnValAsnProGlnAspAsnGlyLysAsnIleSerTrpThpCysGln	193
Db	136953	AACAGAGCTGATGATCTTAATTCGCCAGACAAATGGGAAATGTCTCCAGGGCATGTGCAG	137012
OY	194	ArgSerSerGlnGlnSerIleLysSerLeuAlaTrpArgProArgLysTrpPheCys	213
Db	137013	AAGCTTTCACAGACAGCCCTCCATCCACAGAGCCCTGGAGGCTTAGAGAAATTTGGTTTTGT	137072
OY	214	GlyThnGlyProGlySerLeuLysCysValGlnProArgAspLeuValProCysValPro	233
Db	137073	GGGCGAGGCCCGAGGGTCCCAATGCTGTGTCCACCTTAGGGAGCTTGCTGCTGATCCCA	137132
OY	234	ValAsnSerAlaValAlaSerGluGly-----	242
Db	137133	GCGGCTCTAGCCATGGCTGGAAAGGGGCCCATGTAGATPAGACTCAAGCTGTGGCTTACA	137192
OY	243	AlaSerProLysProTrpGlnLeuProSerGlyValGlnProValGlyAlaLysLysSer	262
Db	137193	GGGTGCCCTTAACCCCTGGCAGATCTTCCACAGTGGTGTTAGACTTTTAGGTACACAGAACTCA	137252
OY	263	ArgTlEGlnValTrpGluProProIleArgPheGlnLysIleTyrGlyAsnProTrpMet	282
Db	137253	AGAAATTGAGGTTTGGAACTTCCCTCCCTGATTTCCAAAGATGTATATGGAAAGGCTTGGATG	137312
OY	283	ProArgGlnLysPheAlaValGlyValGlySerSerTrpArgThpSerAlaArgValVal	302
Db	137313	CCAGGCCAGAAAGTTTGGTCGCAAGGGGAGGAGATCTCATGGAGAAACCTTGTCTTAGGGCACTG	137372
OY	303	GlnLysGlyAsnValGlyTrpGluProProHisArgValProSerGlyAlaProSerSer	322
Db	137373	CAGAAAGGAAATGTGGGGTCCAAAGCCACCAATATGAGAGTCCCTACACGGGCACTTTAGT	137432
OY	323	ArgAlaValAlaArgArgSerProProSerSerArgLeuGlnLysGlyArgSerThpAspSer	342
Db	137433	GAGAGCTGTATAGAAAGGGCCCAATATCCCTTCCACAGACCCCAAGAAATGTATCCACCTTACAG	137492

Oy	343	LeuInhi	sVal	Pro	Lys	Ser	Thr	Asp	Thr	Gln	Gly	Ser	Gln	Pro	Val	Lys	Ala	Ile	Agly	362		
Db	137493	TTTGAC	CGGT	TTGCTT	GGAA	AAAC	CTCC	CAG	CACT	CAAC	CA	CCG	CGG	GA	AG	GA	GC	CA	AG	137552		
Oy	363	Met	Lys	er	Val	Pro	Arg	Lys	Thr	Val	Val	Ala	Glu	Leu	r	Lys	Thr	Val	Ile	Leu	382	
Db	137553	AGG	AAG	CGT	CTT	GGCC	CTT	AC	AA	AGC	CA	AGG	GGT	GG	AGT	TTG	CC	CA	AG	CA	137612	
Oy	383	Leu	Leu	hi	s	Cy	hi	s	A	ple	u	s	A	py	Al	a	r	g	hi	s	402	
Db	137613	CTCT	TTC	AT	C	A	G	A	T	G	A	C	T	G	A	C	T	G	A	C	137672	
Oy	403	Leu	Arg	Phe	Asp	Cys	Pro	Thr	Gly	Phe	Arg	Thr	Y	Met	Gly	Pro	Val	-	Pro	Leu	Cys	422
Db	137673	TTAA	ATTT	AT	TT	AT	TT	AT	TT	AT	TT	AT	TT	AT	TT	AT	TT	AT	TT	AT	137732	
Oy	422	eg	L	g	L	n	P	h	e	P	h	e	P	r	O	p	h	e	G	L	Y	440
Db	137733	TGG	CA	AT	TT	CT	CC	AT	TT	GG	AA	CA	GT	GT	AT	TT	AC	CA	GA	TA	137790	
RESULT	4																					
LOCUS	AC008799																					
DEFINITION	Homo sapiens chromosome 5 clone CTD-2061E19, complete sequence.																					
ACCESSION	AC008799																					
VERSION	AC008799.6																					
KEYWORDS	HTG.																					
SOURCE	Homo sapiens.																					
ORGANISM	Homo sapiens.																					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.																					
AUTHORS	1 (bases 1 to 123098)																					
TITLE	DOE Joint Genome Institute and Stanford Human Genome Center.																					
JOURNAL	Direct Submission																					
REFERENCE	2 (bases 1 to 123098)																					
AUTHORS	DOE Joint Genome Institute.																					
TITLE	Direct Submission																					
JOURNAL	Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA																					
REFERENCE	3 (bases 1 to 123098)																					
AUTHORS	DOE Joint Genome Institute.																					
TITLE	Direct Submission																					
JOURNAL	Submitted (07-MAR-2002) Production Sequencing Facility, DOE Joint																					



Db	54561	GGGCGAAGCCCAAGGTCCTTGGCTGTGTGACAGCTTAGGAGACTTGGTGGCCCTGCATCCCA	54620		
Oy	234	Val-----AsnSerAlaValAlaSerGluGly	242		
Db	54621	GCTGCTCCACCCATGGCTGAAAGGGCCCAACTAGAGCTCGGCTGTGGCTTCAGAAAGT	54680		
Oy	243	AlaSerProIysProIhrProGlnLeuProSerGlyValGluProValGlyAlaLysSer	262		
Db	54681	GCAAGCCCCAAGCGTTGGACGCTCCATGTGGTGTGGACCTGCAGAGTGCAGAGAAGTCA	54740		
Oy	263	ArgIleGluValTrpGluProProIleArgPheGlnLysIleTyrGlyAsnProTrpMet	282		
Db	54741	CAATTTGGGTTTGGGAGACCTCCACCTAGATTCAGATGATGTATGAAATGCCCGGATG	54800		
Oy	283	ProArgGlnLysPheAlaValGlyValGlySerSerTrpArgTrhSerAlaArgVal	302		
Db	54801	ACCGAGCAGAAAGTTTGTCTGCAGAGGGAGCGGCTTCATGTAGAGACCTCTGCTAGGCGAGT	54860		
Oy	303	GlnLysLysAsnValGlyTrpGluProProIhSaIvalProSerGlyAlaProSerSer	322		
Db	54861	CAGAGAGAAAATATAGGGTCAGAGCCACACACAGAGTACCCACTAGACGACACGCTAGT	54920		
Oy	323	ArgAlaValArgArgSerProProSerSerArgLeuGlnLysGlyArgSerTrpAspSer	342		
Db	54921	GGAGCTGTGAGAAAGAGCCACCATCTCCAGACTCCAGAAATGTTAGTCCACCAACAGC	54980		
Oy	343	LeuGlnHisValProGlnLysSerThrAspTrhGlnCysGlnProValLysAlaAlaGly	362		
Db	54981	TTTGACATGTATCACTGGAAAGACGACAGGACATCTATTCAGCCTGTGAAAGCAGCCAG	55040		
Oy	363	MetGlnSerValProTyrLysThrValValAlaGluLeuThrLysThrValGlyIleTyr	382		
Db	55041	AGGAGAGCTCTCCCTCCGCAAGAGTCAACAGGGGCGAAGCTGTCCAGACCTTGGGAGACCTAC	55100		
Oy	383	LeuLeuIscLysHisAspLeuAspValArgHisGlyValLysAspAspHisPheGlyAla	402		
Db	55101	CTCTTGCTCCTACGCGTATCCCGAGTGTGAGACATGTGGAGTCAAGAGATCAATTTTGAAGCT	55160		
Oy	403	LeuArgPheAspCysProTrhGlyPheArgTrhTyrMetGlyProValProLeuCysPhe	422		
Db	55161	TTTAGATTTTACAGCCCTGCTGTGATTTTCAGACGTATGGGGCGTGTAGCCCTTTGCTTT	55220		
Oy	423	GlyGlnPhePheProPheGlyThrAlaValPheTrhGlnCysLeuTyrLeuHisCys	441		
Db	55221	GGCCAAATTTCTCCATTTGGAAACAGCTGATTTACCCTCATCTGCTACCCGATTTGT	55277		
RESULT 6					
AC0044889/c					
LOCUS	AC0044889	189768 bp	DNA linear HTG 24-AUG-2002		
DEFINITION	Homo sapiens chromosome 5 clone RP11-792C1 map 5, WORKING DRAFT				
ACCESSION	AC0044889	SEQUENCE, 36 unordered pieces.			
VERSION	AC0044889.2	GI:8016676			
KEYWORDS	HTG, HTGS_PHASE1, HTGS_DRAFT.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	1 (bases 1 to 189768)				
JOURNAL	Birren, B., Linton, L., Nusbaum, C. and Lander, E.				
REFERENCE	Homo sapiens chromosome 5, clone RP11-792C1				
AUTHORS	Unpublished				
	2 (bases 1 to 189768)				
	Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bede, F.,				
	Boguslavsky, L., Bouknightalter, B., Brown, A., Burkett, G.,				
	Campimino, A., Castle, A., Chepel, Y., Colangelo, M., Collins, S.,				
	Collymore, A., Cooke, P., DeArillano, K., Dewar, K., Diaz, J.S.,				
	Dodgson, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, N., Gage, D.,				
	Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,				
	Grand-Pierre, N., Grant, G., Hagos, B., Heathford, A., Horton, L.,				
	Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,				
	Klein, J., Lakocque, R., Lamazares, R., Landers, T., Lehoczy, J.,				

TITLE  
JOURNAL  
AUTHORS

Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
(Passes 1 to 189768)  
Barron, S., Linton, L., Nusbaum, C., Landre, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bartley, V., Bedalov, F., Boguski, M.S., Borker, A., Brown, A., Burkett, G., Campione, A., Castle, A., Chao, P., Chen, Y., Chinaiolo, M., Collins, S., Collymore, A., Cooke, P., DeArnell, K., Dewar, K., Diaz, J.S., Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Glendy, S., Goette, M., Graham, L., Grand-Pierre, N., Grant, G., Haas, R., Heaford, A., Horton, L., Howard, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatzis, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lebeck, J., Levine, R., Lien, C., Liu, G., Locke, K., MacDonald, P., Margolis, N., McCarthy, M., McEwan, P., McGuck, A., McKernan, K., McQuatters, R., Meldrum, J., Menous, L., Mihova, T., Miranda, C., Miya, Y., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talmas, J., Teste, S., Theodore, J., Tirelli, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 22, 2000 this sequence version replaced gl1:543856.  
All repeats were identified using RepeatMasker:  
Smith, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIRB  
Web site: <http://www.seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
Project Information  
Center project name: L9594  
Center clone name: 792.C.1

Summary Statistics  
Sequencing vector: M13: M7815: 100% of reads  
Chemistry: Dye-Terminalator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 168749 bases at least Q40  
Consensus quality: 180047 bases at least Q30  
Consensus quality: 184014 bases at least Q20  
Insert size: 190000; agarose-fp  
Insert size: 186268; sum-of-contigs  
Quality coverage: 3.8 in Q20 bases; agarose-fp  
Quality coverage: 3.9 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 36 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1015: contig of 1015 bp in length  
\* 1016 1115: gap of 100 bp





```

repeat_region /rpt_family="Alu"
1704. .1759
repeat_region /rpt_family="GA-rich"
1949. .2123
repeat_region /rpt_family="L1"
2652. .3582
repeat_region /rpt_family="L1"
3583. .4287
repeat_region /rpt_family="ERV1"
4295. .5073
repeat_region /rpt_family="L1"
5074. .5105
repeat_region /rpt_family="(CAAA)n"
5106. .5232
repeat_region /rpt_family="L1"
5239. .5956
repeat_region /rpt_family="L1"
8153. .8284
repeat_region /rpt_family="L1"
8335. .8526
repeat_region /rpt_family="L1"
8527. .8644
repeat_region /rpt_family="MaLR"
8633. .8771
repeat_region /rpt_family="MaLR"
8781. .10442
repeat_region /rpt_family="L1"
11035. .11325
repeat_region /rpt_family="Alu"
12515. .12689
repeat_region /rpt_family="MER1_type"
13044. .13144
repeat_region /rpt_family="MIR"
13264. .13307
repeat_region /rpt_family="(CA)n"
13835. .14528
repeat_region /rpt_family="ERVK"
15108. .15333
repeat_region /rpt_family="ERV1"
15588. .15692
repeat_region /rpt_family="MIR"
17652. .17965
repeat_region /rpt_family="Alu"
17966. .18124
repeat_region /rpt_family="(TA)n"
18125. .18164
repeat_region /rpt_family="AT-rich"
18350. .18387
repeat_region /rpt_family="MER1_type"
18388. .18659
repeat_region /rpt_family="Alu"
18660. .18746
repeat_region /rpt_family="MER1_type"
19155. .19257
repeat_region /rpt_family="AchoBo"
19857. .20350
repeat_region /rpt_family="MaLR"
20354. .20447
repeat_region /rpt_family="MIR"
20477. .20680
repeat_region /rpt_family="MIR"
21333. .21520
repeat_region /rpt_family="L1"
22232. .22307
repeat_region /rpt_family="(TAGA)n"
22506. .22720
repeat_region /rpt_family="MER1_type"
22738. .23033
repeat_region /rpt_family="ERV1"
23041. .23231
repeat_region /rpt_family="ERV1"
23330. .23413
repeat_region /rpt_family="MaLR"

```

```

repeat_region 23474. .23666
/rpt_family="MaLR"
repeat_region 23789. .23810
/rpt_family="AT-rich"
repeat_region 24350. .24427
/rpt_family="MER1_type"
repeat_region 25035. .25129
/rpt_family="MIR"
repeat_region 26055. .26135
/rpt_family="L2"
repeat_region 26545. .26767
/rpt_family="MIR"
repeat_region 31261. .31391
/rpt_family="Martiner"
repeat_region 31511. .31567
/rpt_family="CRI"
repeat_region 32245. .32294
/rpt_family="AT-rich"
repeat_region 32743. .32881
/rpt_family="Alu"
repeat_region 33475. .33986
/rpt_family="MER1_type"

```

Alignment Scores:

Prod. No.:	1,12e-55	length:	91323
Score:	1209.00	Matches:	238
Percent Similarity:	75.80%	Conservative:	22
Best Local Similarity:	69.39%	Mismatches:	69
Query Match:	50.04%	Indels:	14
DB:	9	Gaps:	2

US-09-997-610-2\_COPY\_18\_459 (1-442) x AC106901 (1-91323)

```

QY 114 GlnGlnIleSer-LysGlnGlnSerIleGlnGlnValThrTpyAlleuLysAlaPh 133
|||||
Db 72465 GAAGAAATTTCAAAGACGACGAAACATTCAAATTGATTGGGCTGTTAAAGCAT 72524
QY 133 eSerPheIleArgGlnAgluHIsIySerSerGlnuLnleuHIsProAspaInVal11 153
|||||
Db 72525 CAGTTTATTAAGAAACGACGATTAAGTTGAAATTGACGCTGACATGCAT 72584
QY 153 eLysLysLysAsnProPheSerGlnGlnLysPheLysLeuAlaIleGlnIleCysIleCy 173
|||||
Db 72585 AGAAAGAAATTCATTTTGTGAGGAAATTCAGCCACCTACAGCATTTGCATAG 72644
QY 173 sasngLglnLeuAsnValasnpGlnAspAsngLyGlnuLnleuHIsIleSerTrpHrcysG1 193
|||||
Db 72645 TAATGAGAGCTGAATGTATATCCACAAGACATGAGGAAATGTCTCCAGCGCATGCA 72704
QY 193 natrSerSerGlnGlnSerIleLysSerLeuAlaTrpArgProArg---LysTrpPh 212
|||||
Db 72705 AAGCTCTTATGAGCGCTCCATCATCAGCGCTGCGGCTGAGGAAATGATT 72764
QY 212 ecysGlyThngLyProGlySerLeuCyScysValGlnProArgAspLeuValProCysVa 232
|||||
Db 72765 TTGTGGCCGGGCCAGGCTCCCATGCTATGTCAATGAGGATGTGTTCCCTGCGT 72824
QY 232 lPro-----ValasnerAlaValaIleSerG1 241
|||||
Db 72825 CCCAGCTGCTCCACTTGTGATTAATGATGATCAAGTGACGCTACAGGCCATGCTTCAGA 72884
QY 241 uGlyAlaSerProLysProTrpGlnLeuProSerGlyValGlnProValGlyAlaLysLy 261
|||||
Db 72885 GGGTACAAAGCCCAAGCCTTGCGAGCTTCCACATGTGTTGAGCCTGTGGGTACAGAA 72944
QY 261 sSerArgIleGlnValTrpGlnProPheLysPheGlnLysIleTrpGlyAsnProTr 281
|||||
Db 72945 GTTAAGAAATTGAGGTTTGGGAACCTTCGCTAGATTTCAGAGAAATGATGAACACCTG 73004
QY 281 pmeLProArgGlnLysPheAlaValGlyValGlySerSerTrpArgHrSerAlaArgVa 301
|||||
Db 73005 GATGTCAGGACGAGCTTTGCTGTCGAGGACGAGGCGCTCATGAGAAACCTGCTGTAAGGC 73064

```

QY 301 lValGlnLysGlyAsnValGlyTrpGluProHisArgValProSerGlyAlaProSe 321  
Db 73065 AGGAGAGATGGAAATGTGGGTGGACCCCCACACAGAGTACTACTGGGTACTGCC 73124  
QY 321 rSerAGAlaValArgArgSerProProSerSerArgLeuGlnLysGlyArgSerThraS 341  
Db 73125 TACTGAGAGCTGTGAGAGAGAGGACACATCTTCCAGACTCCAGAAATGTTAGATCCACTGA 73184  
QY 341 pSerLeuGlnHisValProGlnLysSerThrAspThrGlnCysGlnProValLysAlaAl 361  
Db 73185 CACGCTTGCCCGCATGTGCGCAAAAGCCACAACAGTCATGTGACGCCCATGAAAGCAGC 73244  
QY 361 asGlyMetGlnSerValProTyrLysThrValValAlaGlnLeuThrLysThrValGlyII 361  
Db 73245 TGGGAGGAGGCGTGTGCAAGCCACAGAGGGGCAACAATGCCCAAGACATGGGAAAC 73304  
QY 381 eTyLeuLeuHisCysHisAspLeuAspValArgHisGlyValLysArgAspHisPheG 401  
Db 73305 CCACGCTTGATGCATGACATGACCTGATGTGAGACATGTGATCAAGACATCATTTTGG 73364  
QY 401 yAlaLeuArgPheAspCysProThrGlyPheArgThrTyrMetGlyProVal-ProLeu 421  
Db 73365 AGCTTTAAGATTGTGACTGCGCTGCTGATTCACAGCTTGATGCGGCGCTGACGCCGCTT 73424  
QY 421 ySPheGlyGlnPhePheProPheGlyThrAlaValAlaPheThrGlnCysLeuTyrLeuHisC 441  
Db 73425 GTTTTGGCCAAATCTCTCCCATTTGGATGGCTGTACTTACCAAAATGCGTGTACCCCAT 73484  
QY 441 ySmet 442  
Db 73485 GTATC 73489  
RESULT 8  
AL139216 112515 bp DNA linear PRI 25-OCT-2001  
LOCUS AL139216  
DEFINITION Human DNA sequence from clone RP5-1017J17 on chromosome  
ACCESSION AL139216 GI:16501095  
VERSION AL139216.14  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 112515)  
AUTHORS Ellington, A.  
TITLE Direct Submission  
JOURNAL Submitted (25-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CH10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
COMMENT  
On Oct 26, 2001 this sequence version replaced gi:15020340.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submissions with  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em: EMBL, Sw:  
SWISSPROT, Tr: TrEMBL, Wp: WORMPEP, Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone configs of human  
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr1

RP5-1017J17 is from the library RPCT-5 constructed by the group of  
Pleter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: PCYPAC2  
This sequence is the entire insert of clone RP5-1017J17 The true  
left end of clone RP5-944N15 is at 108725 in this sequence.

FEATURES  
source  
1..112515  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="p31.2-32.2"  
/clone="RP5-1017J17"  
/clone\_lib="RPCT-5"  
69376..70016  
/note="CpG island"  
/evidence="not\_experimental"  
74591..75390  
/note="CpG island"  
/evidence="not\_experimental"  
misc\_feature  
misc\_feature  
BASE COUNT 33418 a 20758 c 21055 g 37284 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.83e-55 Length: 112515  
Score: 1207.00 Matches: 242  
Percent Similarity: 66.67% Conservative: 34  
Best Local Similarity: 58.45% Mismatches: 83  
Query Match: 49.96% Indels: 55  
DB: 9 Gaps: 4  
US-09-997-610-2\_COPY\_18\_459 (1-442) x AL139216 (1-112515)  
QY 42 SerAlaPheThrValLysLeuSerGlyLysLeuProLeuProPheLysProIleIlePhe 61  
Db 25088 TCTTGTTTATGCAAAAAGATGCGTGCATTTTGCTGCTGCTTGAAGATTTGCTG- 25141  
QY 62 ThrGlyValLeuTyrAsnAlaGlnArgAspLeuLysGlnAlaMetGlyValPheAlaCys 81  
Db 25142 -----GAACTTGAACTTGAGAGAGATTGAGGCTATCTGGCA--- 25177  
QY 82 ArgValProGlyAsnTyrTyrSerSerPheAspValGlnLeuHisCysLysValAsn 101  
Db 25177 ----- 25177  
QY 102 IleTrpLeuMetArgLysGlnIleLeuAlaAsnLysGlnGlnIleSerLysGlnGlnSer 121  
Db 25178 -----GAAAGAAATTTCTAAGCAACGAAAGC 25201  
QY 122 IleGlnGlnValThrTrpValLeuLeuLysAlaPheSerPheIleArgGlnAlaGlnHis 141  
Db 25202 ATTCAGAGGCTGACTTGGGCTGTAAAGCATTCAGATTATTAAGGAGAGAGCAT 25261  
QY 142 LysSerSerLysAsnLeuHisProAspAsnValIleLysLysLysAsnProPheSerGlu 161  
Db 25262 AAAAGATTGGAAAAATTTTACCCTGACATGCAATTAAGAAAAATCCCTTTTTCGAA 25321  
QY 162 GlySPheLysLeuAlaAlaGlnIleCysIleCysAsnGlnGlnLeuAsnValAsnPro 181  
Db 25322 ATGAAATTCAAGCTGGCTGCAGAAATGTGATTAAGTAATGAAAGCCAAATATTAATCAC 25381  
QY 182 GlnAspAsnGlyLysAsnIleSerTrpThrCysGlnArgSerSerGlnGlnSerIleLys 201  
Db 25382 CAAGACAAATGGGAAATATGCTACAGAGCATGTGAGAGCTTCATGGCAGCCCTCGCA 25441  
QY 202 SerLeuAlaTrpArgProArgArg---LysTrpPheCysGlyThrGlyProGlySerLeu 220  
Db 25442 TCACAGCCCTGTGAGGCGCTGTGAGAGATGTGTGTGTGGCTGAGGCCAGCGTCCCG 25501  
QY 221 CysCysValGlnProArgAspLeuValProCysValProValAsnSerAlaVal----- 238  
Db 25502 TGTGTGTGCAAGTGTAGATTTTGTGTGCGCTGCATCCAGCGACTCCAGCGGTGACTAAA 25561

```

QY 239 -----AlaSerGluGlyAlaSerProLysProTyrGln 249
                |||||
Db 25562 AGGATGAAGTACAGCTCAAGCTTACAGAGGCTGAGGCCCAAGCCCTTGACAG 25621
                |||||
QY 250 LeuProSerGlyValGluProValGlyAlaLysSerArgIleGluValTTPGluPro 269
                |||||
Db 25622 CTTCATGTGTGTGTGCTGTGAGTGCACAGATGTCAAGAACTGATGAGAACT 25681
                |||||
QY 270 ProLLeuArgProGlnLysIleTyrGlyAsnProTyrMetProArgGlnLysPheAlaVal 289
                |||||
Db 25682 CCCCCTAGATTTAGAGAGATGTAGTACCCCTGGATGCCACCAAGAGTTGCTGCA 25741
                |||||
QY 290 GlyValGlySerSerTyrArgTyrSerAlaArgValAlaGlnLysGlyAsnValGlyTyr 309
                |||||
Db 25742 GGGGAGAGGCGCATCATGAGAACTGCTGTAGGCAATGCAGAAAGAAATGTGGGTGG 25801
                |||||
QY 310 GluProProHISArgValProSerGlyAlaProSerSerArgAlaValArgArgSerPro 329
                |||||
Db 25802 GAGCCCCCAACAGAGAGCTCTAGTGGGCACTGCTGAGAGCTTGAGAGAGAGCA 25861
                |||||
QY 330 ProSerSerArgLeuGlnLysGlyArgSerThrAspSerLeuGlnHisValProGluLys 349
                |||||
Db 25862 CCATCTCCAGAGCCCAAGAAATGTAGTCCAGTGCACAGCTTGCAACATGTGCCAGAAA 25921
                |||||
QY 350 SerThrAspThrGlnCysGlnProValLysAlaAlaGlyMetGluSerValProTyrLys 369
                |||||
Db 25922 GCCATGAGACACTCAGTCCCAACCCATGAAGCCAGAGGAGGCTATACCTGAAAA 25981
                |||||
QY 370 ThrValValAlaGluLeuThrLysThrValGlyIleTyrLeuLeuHisCysHisAspLeu 389
                |||||
Db 25982 GCCACAGAGGCTGAGCTGTCTCAAGACCATGGAAACCACTTGCTGATCAGTGGAGCCG 26041
                |||||
QY 390 AspValArgHisGlyValLysArgAspHisPheGlyAlaLeuArgPheAspCysProThr 409
                |||||
Db 26042 GATGACAGATTCGAGAGAAAGATCATATTTGACTTTAAATTTGACTGCCCTGCT 26101
                |||||
QY 410 GlyPheArgThrTyrMetGlyProVal-ProLeuCysPheGlyGlnPhePheProPheG1 429
                |||||
Db 26102 GATTTTGGACTTCGATGGGCTGTAGGCCCTTTGTTTGGCCAAATTTCTCCATTGG 26161
                |||||
QY 429 yThrAlaValPheThrGlnCysLeuTyrLeuHisCysMet 442
                |||||
Db 26162 ATCACTGTATTTACCAATGCTGTACCCCATTTGATC 26201
                |||||
RESULT 9
LOCUS HS345P10 137785 bp DNA linear PRI 12-DEC-1999
DEFINITION Human DNA sequence from clone RP3-345P10 on chromosome 22. Contains
a pseudogene similar to 14-3-3 protein theta-subtype, parts of 2
novel genes based on ESTs, a testis-specific mRNA, ESTs, STSs and
GSSs, complete sequence.
ACCESSION 282201
VERSION 282201.1 GI:1843447
KEYWORDS HTG: 14-3-3 protein.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 137785)
AUTHORS Whiteley, M.
TITLE Direct Submission
JOURNAL Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Feb 19, 1997 this sequence version replaced gi:1772951.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 22, constructed by the Sanger Centre Chromosome 22

```

## FEATURES

```

source
1..137785
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="RP3-345P10"
/clone_1ib="RPCI-3"
194..497
/feature="AluSP repeat: matches 1..308 of consensus"
1583..1748
/feature="LINE repeat: matches 5464..5634 of consensus"
1936..2241
/feature="AluSX repeat: matches 10..298 of consensus"
2451..2689
/feature="MIR repeat: matches 10..262 of consensus"
3104..3436
/feature="match: GSS: Em:AQ702221"
4276..4346
/feature="MIR repeat: matches 97..181 of consensus"
4886..4995
/feature="MIR repeat: matches 59..175 of consensus"
5352..6731
/feature="69 copies 20 mer 88 conserved"
join(7248..7402,16817..16920)
/gene="dJ345P10.4"
join(<7248..7402,16817..>16920)
/gene="dJ345P10.4"
/product="dJ345P10.4 (testis-specific mRNA)"
/feature="match: ESTs: Em:AF012350 Em:AF012385"
/evidence="not_experimental"
8997..9586
/gene="dJ345P10.4"
/feature="match: GSS: Em:AQ344456"
9079..9118
/feature="20 copies 2 mer ac 95 conserved"
9137..9384
/feature="HAL1 repeat: matches 698..949 of consensus"
9549..10178
/feature="MER92B repeat: matches 5..636 of consensus"
10282..10637
/feature="MUT1A1 repeat: matches 1..365 of consensus"
10967..13081
/feature="LIMCb repeat: matches 196..2309 of consensus"
13084..13197
/feature="AluJo/FRAM repeat: matches 195..308 of consensus"
13466..13641
/feature="MER5A repeat: matches 1..189 of consensus"
14078..14544
/feature="MLTID repeat: matches 1..501 of consensus"
14674..14815
/feature="MER53 repeat: matches 27..186 of consensus"

```

Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr22>

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information from the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/C.elegans/wormpep> RP3-345P10 is from the library RPCI-3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>

## VECTOR: PCYPAC2

IMPORTANT: This sequence is not the entire insert of clone RP3-345P10. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. The start of this sequence overlaps with sequence AL023801.

## location/Qualifiers

```

1..137785
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="RP3-345P10"
/clone_1ib="RPCI-3"
194..497
/feature="AluSP repeat: matches 1..308 of consensus"
1583..1748
/feature="LINE repeat: matches 5464..5634 of consensus"
1936..2241
/feature="AluSX repeat: matches 10..298 of consensus"
2451..2689
/feature="MIR repeat: matches 10..262 of consensus"
3104..3436
/feature="match: GSS: Em:AQ702221"
4276..4346
/feature="MIR repeat: matches 97..181 of consensus"
4886..4995
/feature="MIR repeat: matches 59..175 of consensus"
5352..6731
/feature="69 copies 20 mer 88 conserved"
join(7248..7402,16817..16920)
/gene="dJ345P10.4"
join(<7248..7402,16817..>16920)
/gene="dJ345P10.4"
/product="dJ345P10.4 (testis-specific mRNA)"
/feature="match: ESTs: Em:AF012350 Em:AF012385"
/evidence="not_experimental"
8997..9586
/gene="dJ345P10.4"
/feature="match: GSS: Em:AQ344456"
9079..9118
/feature="20 copies 2 mer ac 95 conserved"
9137..9384
/feature="HAL1 repeat: matches 698..949 of consensus"
9549..10178
/feature="MER92B repeat: matches 5..636 of consensus"
10282..10637
/feature="MUT1A1 repeat: matches 1..365 of consensus"
10967..13081
/feature="LIMCb repeat: matches 196..2309 of consensus"
13084..13197
/feature="AluJo/FRAM repeat: matches 195..308 of consensus"
13466..13641
/feature="MER5A repeat: matches 1..189 of consensus"
14078..14544
/feature="MLTID repeat: matches 1..501 of consensus"
14674..14815
/feature="MER53 repeat: matches 27..186 of consensus"

```

repeat_region	15110..15434 /note="HAL1 repeat: matches 155..499 of consensus"
repeat_region	15448..15632 /note="HAL1 repeat: matches 1094..1296 of consensus"
repeat_region	15772..15873 /note="HAL1 repeat: matches 1431..1533 of consensus"
repeat_region	15878..15988 /note="AluY repeat: matches 1..111 of consensus"
repeat_region	15997..16043 /note="Alu repeat: matches 255..301 of consensus"
repeat_region	16044..16067 /note="I2 copies 2 mer aa 95 conserved"
repeat_region	16249..16563 /note="AluXx repeat: matches 1..309 of consensus"
repeat_region	17306..17475 /note="L2 repeat: matches 2527..2710 of consensus"
repeat_region	17821..17886 /note="I3 copies 2 mer tt 81 conserved"
repeat_region	17832..17891 /note="3 copies 20 mer 85 conserved"
repeat_region	17919..18229 /note="AluSg repeat: matches 1..310 of consensus"
misc_feature	19712..20092 /note="match: GSS: Em:AQ166214"
repeat_region	20737..21046 /note="AluSp repeat: matches 1..311 of consensus"
repeat_region	21079..24772 /note="LIM4 repeat: matches 2179..5810 of consensus"
repeat_region	24773..25081 /note="Aluub repeat: matches 1..309 of consensus"
repeat_region	25082..25525 /note="LIM4 repeat: matches 5810..6286 of consensus"
repeat_region	25545..25705 /note="LIM5 repeat: matches 6128..6295 of consensus"
repeat_region	26004..26282 /note="AluSX repeat: matches 34..312 of consensus"
repeat_region	26410..26509 /note="L2 repeat: matches 2583..2709 of consensus"
gene	complement(join(26812..27006,35225..35343,42569..42665)) /gene="d345P10.2"
CDS	complement(join(<26812..27006,35225..35343,42569..>42665))) /gene="d345P10.2"
	/note="match: ESTs: Em:AA497915" /codon_start=2 /evidence=not_experimental /product="dJ345P10.2 (part of a novel gene)" /protein_id="CAB63029.1" /db_xref="GI:65722120"
repeat_region	/translation="PRQVAVDSLEACDQHGYLVYTKAKRWSDLSKNLEFDNENGILRRRDIKMLVGGDIPLRPREFRKIMARVPTDEGSHITVOEFLQIKGINSPVAHRCALDDYFNMGHFTRPOOLEMKELQSTERAVALA" /size..27206 /note="MER5A repeat: matches 9..93 of consensus"
repeat_region	27395..27705 /note="AluY repeat: matches 1..311 of consensus"
repeat_region	30575..30981 /note="MSTA repeat: matches 1..426 of consensus"
misc_feature	complement(31305..31756) /gene="dJ345P10.2"
misc_feature	/note="match: GSS: Em:AQ138484" complement(31307..31543) /gene="dJ345P10.2"
misc_feature	/note="match: GSS: Em:AQ269272" 31874..32097
repeat_region	/note="LTR8 repeat: matches 1..227 of consensus"
repeat_region	32106..32395 /note="AluSc repeat: matches 1..291 of consensus"
repeat_region	32396..32544 /note="LIMB3 repeat: matches 6028..6181 of consensus"
repeat_region	33191..33417 /note="MIR repeat: matches 13..249 of consensus"
repeat_region	33588..33596 /note="MIR repeat: matches 157..222 of consensus"

repeat_region	35927..36443	/note="L1M4 repeat: matches 3997..4541 of consensus"
repeat_region	36680..36998	/note="L1M6 repeat: matches 4886..5232 of consensus"
misc_feature	36878..37407	/note="match: GSS: Em:AD472723"
repeat_region	37100..37275	/note="L1M6 repeat: matches 5329..5500 of consensus"
repeat_region	37299..37353	/note="MER58 repeat: matches 2366..2420 of consensus"
repeat_region	37542..37754	/note="L1M6 repeat: matches 267..482 of consensus"
repeat_region	37761..37882	/note="L1M4 repeat: matches 3598..3718 of consensus"
repeat_region	37888..37948	/note="L1 repeat: matches 4556..4616 of consensus"
repeat_region	37952..38257	/note="L1P1A12 repeat: matches 4351..4642 of consensus"
repeat_region	38258..38364	/note="A1u6 repeat: matches 1..306 of consensus"
repeat_region	38565..39352	/note="L1P1A12 repeat: matches 4642..5509 of consensus"
repeat_region	39353..39666	/note="A1u5 repeat: matches 1..308 of consensus"
repeat_region	39667..39870	/note="L1P1A12 repeat: matches 5509..5702 of consensus"
repeat_region	40035..40516	
Alignment Scores:		
Pred. No.:	2.97e-55	Length: 137785
Score:	1205..00	Matches: 249
Percent Similarity:	71.86%	Conservative: 37
Best local Similarity:	62.56%	Mismatches: 88
Query Match:	49.88%	Indels: 24
DB:	9	Gaps: 5
US-09-997-610-2_COPY_18_459 (1-442) x HS345P10 (1-137785)		
Oy	65	LeuTyrAsnIaGInIrgAspLeuYsgIuIaMeTgIyAlaPheAla-----Cys 81
Db	77495	CTGAGGGGGCTCAGATGAGATGAGCGACCTTTGTTGGAACTGAGCAAAAGCGACTTCT
Oy	82	ArgValProGlyAsnTyrTyrSerSerPheAspValGluLeuHisHsIcysLysValAsn 101
Db	77555	TACATTTTTCAGAAAAGACTGCGACACTTTTGGCC-----TCGCTTAAGAT
Oy	102	ILeIrr-----LeuMetArgLysGInIleLeuAlaAsnLysGInuIleSerLys 118
Db	77603	TTTGGAACCTTGTACACTGTAGAGACATGATATTATGGGTACTGTGGCAGAAATTTCTAGG
Oy	119	GInoInserIleGInGInuValThrTrpValLeuLeuLysAlaPheSerPheIleArgIu 138
Db	77663	CAGCAAAAGCTTTAAAGAGTGACTTGGGTGATGTTAAAGCATTTCACTTTTATGAGGGA
Oy	139	AlaGluHisLysSerSerGluAsnLeuHisProAspAsnValIleLysLysLysAsnPro 158
Db	77723	GCAAGATATAAACTTCAGAAAATTTTGGACGCTGCATGTGGATGAGAAAAGAAATCCCA
Oy	159	PheSerGInuGlyLysPheLysLeuAlaGluIleCysIleLysAsnGInuGluLeuAsn 178
Db	77783	TTTTCTGAGCAGAAATTCAGACTGGGTGGCAGAAATGTGCATTAAGTAAAGCAGAGGAAT
Oy	179	ValAsnProGInAspAsnGluLysuIleSerTrpThrCysGInAArgSerSerGInu 198
Db	77843	ATTATATCCCAACACATGTAATAATGTCTCCACGAGCATGTCAAGACTTCTACAGCAG
Oy	199	SerLeuSerLeuLeuIleArgProArgTrg---LysTrpPheCysGInTrgIlePro 217
Db	77993	CCCTCCCATCAAGAGCTCGAGAGCTGAGAGGAGAAAAAGCAGTTTCATAGAGCGGCCCA
Oy	218	GlySerLeuCysCysValGInProArgAspLeuValProCysValProValAsnSer--- 236
Db	77963	GGGTCCTGAGCTGTGTGAGACTAGGAGCTTGTGGCTTCGATCCACAGAGCTCCAGGC

OY 237 -----AlaValAlaSerGluGlyAlaSerProLys 246  
Db 78023 ATGGCTGAAGAGGCCAATTAGAGCTTGGGCGCATGCTTCAGAGGAGGACCAAGCCTTCAAG 78082  
OY 247 ProTPrGlnLeuProSerGlyValGluProValGlyAlaLysSerSerArgIleGluVal 266  
Db 78083 CCTTGACAGCTTCACATGATGTTGAGCCCTGTAGAGCCAGACAGAGTCAACAAATTGAGGTT 78142  
OY 267 TTPGLeuProProIleArgPheGlnLysIleTyrGlyAsnProTPrMetProArgGlnLys 286  
Db 78143 TGGGAACCTTCCTGATTTCAAGATCTACGGAACCCGTGATGTCTAAGACAGAG 78202  
OY 287 PheAla-ValGlyValGlySerSerTrrPargTrrSerAlaArgValAlaGlnLysGlyAs 306  
Db 78203 TTTGGCCGACAGGGGCGGCTTCATPAGACAACTTGTGAGGAGTGCAGAGGAGAAA 78262  
OY 306 nValGlyTrrPgluProProHlaArgValProSerGlyAlaProSerSerArgAlaValAr 326  
Db 78263 TGTGGGGGACAGGGGCCCCACACAGAGTCCCTACTGGGCGACTGCTAGTGCAGCTGTGAG 78322  
OY 326 gArgSerProProSerSerArgLeuGlnLysGlyArgSerThrAspSerLeuGlnHisVa 346  
Db 78323 AAGAGGGCCACCATCTCCAAACCCAGAGGAGTACATCCATGAAAGCTTGACACCTGC 78382  
OY 346 lProGluLysSerThrAspThrGlnGlyGlnProValLysAlaAlaGlyMetLysSerVa 366  
Db 78383 TCCTGGAAGAAGCTGCAGACACTCAATGATGCCCATGAAGACGCGGAGGAGAGCTGT 78442  
OY 366 lProTrrLysThrValAlaValAlaGluLeuThrLysThrValGlyIleTyrLeuLeuHisCy 386  
Db 78443 ACCCTGAAGAGCCACAGGGGCGGAGGCTGCCCAAGACCATGGGAACCCACTTTCATCA 78502  
OY 386 shLAspLeuAspValAlaArgHisGlyValLysArgAspHisPheGlyAlaLeuArgPheAs 406  
Db 78503 GTGTGACCTGATGATGACATGAGATCAAGAGATCATTTAGAGCTTTAAGATTGGA 78562  
OY 406 pCysProThrGlyPheArgThrTyrMetGlyProVal-ProLeuGlyPheGlyGlnPheP 426  
Db 78563 CTGCCCCACTGGATTCAGCTTCGATGAGGCGCTGTAGCCCATTTGTTGGCCAAATTTC 78622  
OY 426 hepProheGlyThrAlaValPheThrGlnCysLeuTyrLeuHisCysMet 442  
Db 78623 TCCCATTTTGAATGCGTGTATTACCAACACCTGTACCTCATTTGATC 78672  
RESULT 10  
AC009047 168080 bp DNA linear HTG 21-JUN-2000  
LOCUS AC009047/c  
DEFINITION Homo sapiens chromosome 16 clone RP11-204G18, WORKING DRAFT  
SEQUENCE, 37 unordered pieces.  
AC009047  
VERSION AC009047.5 GI:8575942  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
AUTHORS 1 (bases 1 to 168080)  
TITLE DOE Joint Genome Institute.  
JOURNAL Sequencing of Human Chromosome 16  
AUTHORS Unpublished  
REFERENCE 2 (bases 1 to 168080)  
TITLE DOE Joint Genome Institute.  
JOURNAL Direct Submission  
AUTHORS Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information

Center Project Name: 484863  
Center Clone Name: RP01-11\_204G18  
-----  
Summary Statistics  
Consensus quality: 140999 bases at least Q40  
Consensus quality: 153370 bases at least Q30  
Consensus quality: 158861 bases at least Q20  
Estimated insert size: 161080; agarose-fp estimation  
Estimated insert size: 164480; sum-of-ctnigs estimation  
Quality coverage: 3.7 in Q20 bases; agarose-fp estimation  
Quality coverage: 3.62 in Q20 bases; sum-of-ctnigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 37 ctnigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the ctnigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 1028: contig of 1028 bp in length  
\* 1029: gap of unknown length  
\* 1129: contig of 1184 bp in length  
\* 2313: gap of unknown length  
\* 2412: contig of 1140 bp in length  
\* 3552: gap of unknown length  
\* 3553: gap of unknown length  
\* 3652: contig of 1562 bp in length  
\* 3653: gap of unknown length  
\* 5215: gap of 1209 bp in length  
\* 5315: contig of 1209 bp in length  
\* 6523: gap of unknown length  
\* 6623: contig of 1104 bp in length  
\* 6624: gap of unknown length  
\* 7727: gap of 1200 bp in length  
\* 7828: gap of unknown length  
\* 9028: gap of unknown length  
\* 9127: contig of 1283 bp in length  
\* 9128: gap of unknown length  
\* 10411: contig of 2189 bp in length  
\* 10511: gap of unknown length  
\* 12700: contig of 2111 bp in length  
\* 12800: gap of unknown length  
\* 14811: contig of 2840 bp in length  
\* 14910: gap of unknown length  
\* 15010: contig of 2789 bp in length  
\* 17851: gap of unknown length  
\* 17950: contig of 2495 bp in length  
\* 20740: gap of unknown length  
\* 20840: gap of unknown length  
\* 23335: gap of 3148 bp in length  
\* 23435: gap of unknown length  
\* 26582: contig of 3499 bp in length  
\* 26683: gap of unknown length  
\* 30181: contig of 2763 bp in length  
\* 30282: gap of unknown length  
\* 33044: gap of unknown length  
\* 33045: gap of unknown length  
\* 33145: contig of 4258 bp in length  
\* 37403: gap of unknown length  
\* 37503: contig of 3119 bp in length  
\* 40622: gap of unknown length  
\* 40721: gap of 4717 bp in length  
\* 45438: gap of unknown length  
\* 45439: gap of unknown length  
\* 45539: contig of 3724 bp in length  
\* 49263: gap of unknown length  
\* 49363: gap of 3623 bp in length  
\* 52985: gap of unknown length  
\* 52986: gap of unknown length  
\* 53086: gap of 4082 bp in length  
\* 57167: gap of unknown length  
\* 57267: gap of unknown length  
\* 60143: contig of 2876 bp in length  
\* 60243: gap of unknown length  
\* 60244: gap of unknown length  
\* 64484: gap of 4240 bp in length  
\* 64584: gap of unknown length  
\* 68613: contig of 4030 bp in length  
\* 68614: gap of unknown length  
\* 68713: gap of unknown length  
\* 74972: contig of 6259 bp in length  
\* 74973: gap of unknown length  
\* 75073: contig of 3613 bp in length  
\* 78685: gap of unknown length  
\* 78785: gap of unknown length





clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatem, M., Catanesi, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://pacpac.med.buffalo.edu>)

#### VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is Rp11-318p5, 2000 bp overlap; the clone sequenced to the right is Rp11-299014, 2000 bp overlap. Actual start of this clone is at base position 153702 of Rp11-318p5; actual end is at base position 4782 of Rp11-299014.

Data from AC016715 was used to finish this clone, AC011238.

Polymorphisms have been identified between AC016715 and AC011238.

#### Location/Qualifiers

1. 155764

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="2"

/map="2"

/clone="RP11-374C5"

/clone\_11b="RPCI-11"

195. 589

/rpt\_family="L2"

repeat\_region

misc\_feature

/note="match to EST A1471723 (NID:g4333813) ta16a05.x1"

misc\_feature

/note="match to EST AA443394 (NID:g2156069) zw87b01.r1"

misc\_feature

/note="match to EST AW952412 (NID:g8142094)"

misc\_feature

/note="match to EST BE856700 (NID:g10369992)"

misc\_feature

/note="similar to Mus musculus EST BG094183 (NID:g12576746) uw72f01.x1"

misc\_feature

/note="similar to Rattus norvegicus EST B1289336 (NID:g14946795)"

misc\_feature

/note="match to EST AA328123 (NID:g1980388)"

misc\_feature

/note="match to EST A1471723 (NID:g4333813) ta16a05.x1"

misc\_feature

/note="match to EST BE466172 (NID:g9511947) hz60c11.x1"

misc\_feature

/note="match to EST N39733 (NID:g1163278) yx92g01.r1"

misc\_feature

/note="match to EST R51616 (NID:g813518) y977b04.s1"

repeat\_region

misc\_feature

/rpt\_family="Alu"

misc\_feature

/note="match to EST AA625701 (NID:g2538088) zu86c07.s1"

misc\_feature

repeat\_region

misc\_feature

misc\_feature

misc\_feature

repeat\_region

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

2644. 2671

/rpt\_family="(T)n"

2659. 3087

/note="match to EST AL603583 (NID:g15167089)"

3474. 4110

/note="match to EST BF348692 (NID:g11283118)"

3958. 4293

/note="similar to Rattus norvegicus EST A1058342 (NID:g3332119)"

4046. 4311

/note="match to EST AW952412 (NID:g8142094)"

4046. 4305

/note="match to EST A1471723 (NID:g4333813) ta16a05.x1"

4046. 4293

/note="similar to Rattus norvegicus EST B1289336 (NID:g14946795)"

4046. 4247

/note="similar to Mus musculus EST BG094183 (NID:g12576746) uw72f01.x1"

4049. 4305

/note="match to EST AA625701 (NID:g2538088) zu86c07.s1"

4377. 4551

/rpt\_family="MER2\_type"

4885. 5179

/rpt\_family="Alu"

5612. 5735

/rpt\_family="Alu"

5865. 6033

/rpt\_family="CR1"

6109. 6113

/note="match to EST AV725468 (NID:g10830886)"

6687. 6729

/rpt\_family="AT\_rich"

6942. 7082

/note="match to EST A1907606 (NID:g6498041)"

7227. 7535

/rpt\_family="Alu"

7230. 7257

/rpt\_family="(TTTG)n"

7543. 7609

/rpt\_family="MER1\_type"

7963. 7990

/rpt\_family="AT\_rich"

8501. 8549

/note="match to EST N39733 (NID:g1163278) yx92g01.r1"

8961. 9280

/note="match to EST AA883177 (NID:g2994653) aml18a07.s1"

8962. 9229

/note="match to EST BF837074 (NID:g12189239)"

9359. 9680

/rpt\_family="Alu"

9656. 9676

/rpt\_family="AT\_rich"

10114. 10445

/note="match to EST W60003 (NID:g1366762) zc92g08.s1"

10382. 10435

/rpt\_family="MIR"

10562. 10566

/note="match to EST AA446382 (NID:g2159047) zw58c01.r1"

10678. 10980

/rpt\_family="Alu"

10941. 11188

/note="match to EST AV725468 (NID:g10830886)"

10973. 11380

/note="match to EST AA446382 (NID:g2159047) zw58c01.r1"

11112. 11453

/note="match to EST BE856700 (NID:g10369992)"

11112. 11425

/note="match to EST N39733 (NID:g1163278) yx92g01.r1"

11117. 11142

/note="match to EST AA443394 (NID:g2156069) zw87b01.r1"

11155. 11792

/note="match to EST N26638 (NID:g1140986) yx92g01.s1"

```

misc_feature      11243..11420
                    /note="match to EST AV725468 (NID:910830886)"
misc_feature      11384..12306
                    /note="match to EST AL535373 (NID:912798866)"
repeat_region     11581..11621
                    /rpt_family="L1"
repeat_region     11654..11690
                    /rpt_family="(TTTTC)n"
repeat_region     11795..12448
                    /rpt_family="L1"
misc_feature      12696..13327
                    /note="match to EST AL565893 (NID:912917719)"

Alignment Scores:
Pred. No.:        6.83e-55      Length:      155764
Score:            1199.50       Matches:     249
Percent Similarity: 74.18%      Conservative: 21
Best Local Similarity: 68.41%   Mismatches: 73
Query Match:      49.65%       Indels:      22
DB:               9            Gaps:           4

US-09-997-610-2_COPY_18_459 (1-442) x AC011238 (1-155764)

QY 98 CysLysValAsnIleTrpLeuMetArgLysGlnIle-----LeuAlaAsn 112
    |||  :::::|||||  :::::|||||  |||||
Db 29905 TGCCTTAAGATTGTTGGAACTTGAACCTTGAATAGATGATTAGGGTATCTGGCA--- 29849

QY 113 LysGluGluIleSerLysGlnGlnSerIleGlnGluValThrTrpValLeuLeuLysAla 132
    |||  :::::|||||  :::::|||||  :::::|||||  :::::|||||  :::::|||||
Db 29848 ---GAAGAAATGCTTAAACAGCAACCAAGCAATTCAGAGTGTGATGCTGTGAGGGCA 29792

QY 133 PheSerPheIleArgGluLagLunHisLysSerSerGluAsnLeuHisProAspAsnVal 152
    |||  :::::|||||  :::::|||||  :::::|||||  :::::|||||  :::::|||||
Db 29791 TTCAGTTTATTAAGGAGACAGACGATTAAGTTCCAGAAACTTGACGCTGACAAAGCA 29732

QY 153 IleLysLysLysAsn-ProPheSerGluGluLysPheLysLeuAlaAlaGluIleCys11 172
    |||  :::::|||||  :::::|||||  :::::|||||  :::::|||||  :::::|||||
Db 29731 ATGAAAGAAAGAAATCCCATTTCTGAGAGAAATCCAAAGCCAGCCTCAAGAAATTTGCAT 29672

QY 172 eCysAsnGluGluLeuAsnValAsnProGlnAspAsnGlyGluAsnIleSerTrpIrcy 192
    |||  :::::|||||  :::::|||||  :::::|||||  :::::|||||  :::::|||||
Db 29671 AAGTAATTTGGGCCCAAAATTTAATCCCAAGCAATGGGAAATGCTCCAGGCGACG 29612

QY 192 sGlnArgSerSerGlnGlnSerIleLysSerLeuAlaIleTrpArgProArgArgLys---Tr 211
    |||  :::::|||||  :::::|||||  :::::|||||  :::::|||||  :::::|||||
Db 29611 TCGAGAGTCTTCATGCGAGCCCTCTCATCACAGGCTCGAGGCGTGAAGGAAAGAGTG 29552

QY 211 pPheCysGlyThrGlyProGlySerLeuLysCysValGlnProArgAspLeuValProcy 231
    |||  :::::|||||  :::::|||||  :::::|||||  :::::|||||  :::::|||||
Db 29551 GTTTCATGGGCTGGGCTCCAGGGTCCCATGCTGTACAGTCTAAGGAGACTTTGTGCCCTG 29492

QY 231 sValProValAsnSerAla-----ValAlaSer 240
    |||  :::::|||||  :::::|||||  :::::|||||  :::::|||||  :::::|||||
Db 29491 CATCCAGCTGCTCCAGCCAGCGCTGAAGGGGCCAATGAGAGCTCAGACCTGTGGCTTC 29432

QY 240 rGluGluAlaSerProLysProTrpGlnLeuProSerGlyValGluProValGlyAlaLy 260
    |||  :::::|||||  :::::|||||  :::::|||||  :::::|||||  :::::|||||
Db 29431 AGAGGATGCAAGCCTCAACACTTGGCAGCTTCACGCTGGTGGCGCTGTGAGTGCATA 29372

QY 260 sLysSerArgIleGluValTrpGluProProIleArgPheGlnLysIleTrpGlyAsnPr 280
    |||  :::::|||||  :::::|||||  :::::|||||  :::::|||||  :::::|||||
Db 29371 AATTCAGAAATTTGAGGTTGGGAACCTCCACCTACATTTTCAGAAAGACGATGGAATGC 29312

QY 280 oTrpMetProArgGluLysPheAlaValGlyLysSerTrpArgThrSerAlaAr 300
    |||  :::::|||||  :::::|||||  :::::|||||  :::::|||||  :::::|||||
Db 29311 CCGGATGCCAGGACAGAAATTTGCTAGGGGGTGGGCTCTCATGTGGAACCTCTGCTAG 29252

QY 300 gValValGlnLysGlyAsnValGlyTrpGluProProHisArgValProSerGlyAlaPr 320
    |||  :::::|||||  :::::|||||  :::::|||||  :::::|||||  :::::|||||
Db 29251 GGCAGTGCAGAAAGGAAATGTTGGGTGGAGCCGCCACAGAGTCCCTACTGCGGACACC 29192

QY 320 oSerSerArgAlaValArgArgSerProProSerSerArgLeuGlnLysGlyArgSerTh 340
    |||  :::::|||||  :::::|||||  :::::|||||  :::::|||||  :::::|||||

```

```

Db 29191 ACCTAGTAAGCTGTGAGAAATAGGGCCATCATCTCCAGACTCCAGAAATGATGATCCAC 29132
QY 340 rAspSerLeuGlnHisValProGluLysSerTrpAspThrGlnCysLlnProValLysAl 360
    |||  :::::|||||  :::::|||||  :::::|||||  :::::|||||  :::::|||||
Db 29131 TGACAGCTGTACCATGCGCTGAGAAAGCCTCAGACATCAACA-CCGCGAGGAAGGT 29073

QY 360 aAlaGlyMetClnSerValProTrpLysTrpValValAlaGluLeuThrLysThrValG1 380
    |||  :::::|||||  :::::|||||  :::::|||||  :::::|||||  :::::|||||
Db 29072 AGCTGGAGAGGAGGCTTACCCTCGCAAGGCACAGGGGTGGGGCTGCCCAAGACTATGGG 29013

QY 380 yIleTrpLeuLeuHisCysHisAspLeuAspValArgHisGlyValLysArgAspHisPh 400
    |||  :::::|||||  :::::|||||  :::::|||||  :::::|||||  :::::|||||
Db 29012 AACCTGCCTTCTGCATCAGCATGACCTGCATGATGAGACATGAGATCAAAAGACATCATTT 28953

QY 400 eGlyAlaLeuArgPheAspProThrGlyPheArgThrTrpMetGlyProVal-ProL 420
    |||  :::::|||||  :::::|||||  :::::|||||  :::::|||||  :::::|||||
Db 28952 TGGAGCTTAAAGATTGACATGCTCCACTGGGTTTCACACTTGATGGAGGGGCTGTAGCCC 28893

QY 420 eucysPheGlyGlnPhePheProPheGlyThrAlaValPheThrGlnCysLeuTrpLeuH 440
    |||  :::::|||||  :::::|||||  :::::|||||  :::::|||||  :::::|||||
Db 28892 TTTGTTTGGCCCAATTTCTCCATTTAGATTGGCTGTATTTACCACCAATGCTGTACCCCT 28833

QY 440 lscysMet 442
    |||  :::::|||||  :::::|||||  :::::|||||  :::::|||||  :::::|||||
Db 28832 ATTGTATC 28825

RESULT 13
AC016715/c
LOCUS
DEFINITION
AC016715
AC016715 Homo sapiens chromosome 2 clone RP11-311D8, WORKING DRAFT SEQUENCE,
9 unordered pieces.
AC016715
AC016715.4 GI:8572513
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
Homo sapiens.
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 187898)
Waterston,R.H.
AUTHORS
The sequence of Homo sapiens clone
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 187898)
AUTHORS
Waterston,R.H.
TITLE
Direct Submission
JOURNAL
Submitted (04-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
On Jun 19, 2000 this sequence version replaced gl:8569912.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- project information -----
Center project name: H_NH0311D08
----- Summary Statistics -----
Sequencing vector: M13; 80%
Chemistry: Dye-primer ET; 80% of reads
Chemistry: Dye-terminator Big Dye; 20% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 181767 bases at least Q40
Consensus quality: 183800 bases at least Q30
Consensus quality: 185054 bases at least Q20
Insert size: 195000; agarose-fp
Insert size: 187098; sum-of-contigs
Quality coverage: 5.73 in Q20 bases; agarose-fp
Quality coverage: 5.24 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```







GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 19, 2003, 14:53:22 ; Search time 217.528 Seconds

(without alignments)  
2084.671 Million cell updates/sec

Title: US-09-997-610-2\_COPY\_18\_45

Perfect score: 162

Sequence: 1 AGPRAHRRPREVGRGAPGLPQYIGE1 28

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DEV=x1h  
-O=/cgn2.1/USPTO/US09997610/runat\_10022003.160823.23843/app\_query.fasta.1.1635  
-DB=EST -QFMT=fastap -SUFFIX=1st -MINMATCH=0.1 -LOOPEXT=0 -LIST=45  
-UNITS=bits -START=1 -END=1 -MATRIX=blonsun62 -TRANS=human40.cdi -ALIGN=15 -MODE=LOCAL  
-DOCALLIG=200 -THR SCORE=100 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09997610.ecgn.1.1.1871.etunat\_10022003.160823.23843 -NCPU=6 -ICPU=3  
-NARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estmu:\*  
5: em\_estnu:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vln:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	56.8	372	10	BB772103
2	88	54.3	447	9	AA934284
3	88	54.3	458	9	AA934212
4	88	54.3	531	9	AA934261
5	86	53.1	399	10	BE588347
6	86	53.1	445	9	AA857968
7	86	53.1	499	9	BE483103
8	86	53.1	501	9	AA847768
9	86	53.1	524	12	BE846038
10	86	53.1	577	10	AA597164
11	86	53.1	662	10	AA257669
12	85.5	52.8	609	10	AA675980
13	85.5	52.8	626	10	AA347951
14	85	52.5	391	14	H30955
15	85	52.5	490	9	AA585547
16	85	52.5	491	9	AA585556
17	85	52.5	563	10	AA348031
18	85	52.5	570	10	AA347952
19	85	52.5	571	10	AA225414
20	85	52.5	588	9	AA933455
21	85	52.5	598	10	AA675844
22	85	52.5	600	10	AA409438
23	85	52.5	637	10	AA682789
24	85	52.5	645	10	AA409436
25	85	52.5	645	10	AA409452
26	85	52.5	647	10	AA409421
27	85	52.5	650	10	AA409405
28	85	52.5	652	10	AA675873
29	85	52.5	656	10	AA497698
30	85	52.5	656	10	AA348029
31	85	52.5	657	10	AA225347
32	84	51.9	457	12	BE639395
33	84	51.9	565	9	A1515686
34	84	51.9	1101	17	CNE0024A
35	83	51.2	358	10	BE113822
36	83	51.2	388	9	A1444918
37	83	51.2	521	14	BQ266989
38	83	51.2	573	13	BE923552
39	83	51.2	599	10	BE306946
40	83	51.2	601	14	BQ266473
41	83	51.2	603	10	BE288346
42	83	51.2	666	14	BQ442816
43	83	51.2	674	14	BQ442305
44	83	51.2	726	14	BQ771348
45	83	51.2	738	12	BE864336

#### ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
BB772103	BB772103	372 bp mRNA	BB772103	BB772103	GI:16213794	house mouse.	Mus musculus	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. (bases 1 to 372)	Akimura,T., Arikawa,T., Carninci,P., Furuno,M., Hanaoka,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii



QY 21 LeuProGIntyThrglygu 27  
|||||  
Db 364 TTACCCGCAACCTGATAA 384

RESULT 3  
LOCUS  
DEFINITION  
AA934212 458 bp mRNA linear EST 26-MAR-1999  
SMOVL3CAN09C06 Onchocerca volvulus infective larva cDNA  
(SAM94WL-OVL3) Onchocerca volvulus cDNA clone onch318 5' similar to  
WP:C29F4.1 CE03038 COLLAGEN ; contains element PIR5 repetitive  
element ;, mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AA934212  
AA934212.1 GI:3091369  
EST.  
Onchocerca volvulus.  
Onchocerca volvulus.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
Onchocercidae; Onchocerca.  
1 (bases 1 to 458)  
Williams, S.A., Lizotte-Waniewski, M., Laney, S., Wenhong, L., Hillier,  
L., Allen, M., Bowles, L., Geisel, S., Jost, S., Kucaba, T., Martin, J.,  
Stephens, M., Theising, B., White, Y., Wylie, T., Chappell, J., Person, B.,  
Gibbons, M., Harvey, N., Pape, D., Chamberlain, A., Morales, R.,  
Schurk, R., Ritter, E., Kohn, S., Underwood, K. and Marra, M.  
Molecular Parasitology OVL3  
Unpublished (1998)  
Contact: Steven A. Williams  
Molecular Parasitology  
Smith College Department of Biological Sciences  
Department of Biological Sciences, Clark Science Center, Smith  
College, Northampton, MA, 01063, USA  
Tel: 4135853826  
Fax: 4135853786  
Email: genome@smith.edu  
The library was constructed by Wenhong Lu. The library is available  
from Dr. S.A. Williams, email genome@smith.edu When requesting this  
clone from Dr. Williams, please reference the Williams lab clone id  
- SMOVL3CAN09C06  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 456.  
Location/Qualifiers

FEATURES  
source  
1..458  
/organism="Onchocerca volvulus"  
/strain="Sierra Leone"  
/db\_xref="taxon:6282"  
/clone="Onch318"  
/clone\_11b="Onchocerca volvulus infective larva cDNA  
(SAM94WL-OVL3)"  
/lab\_host="XLI-Blue MRF"  
/note="Vector: lambda Unizap XR; Site\_1: Ecor I; Site\_2:  
Xho I; Cutaneous filarial nematode parasite of humans.  
mRNA was prepared from third stage infective larvae of  
Onchocerca volvulus isolated from mosquitoes 10 days after  
infection and converted to double stranded cDNA using  
reverse transcriptase and oligo(dT) followed by RNase H  
and DNAPol I. The library had 1.8 x 10<sup>5</sup> independent  
recombinants and average insert size was 900 base pairs.  
The library was constructed by Wenhong Lu. The library is  
available from Dr. S.A. Williams, email genome@smith.edu."

BASE COUNT 113 a 122 c 121 g 102 t

ORIGIN

Alignment Scores:  
Pred. No.: 40.5 Length: 458  
Score: 88.00 Matches: 16  
Percent Similarity: 66.67% Conservative: 2  
Best Local Similarity: 59.26% Mismatches: 9  
Query Match: 54.32% Indels: 0  
Gaps: 0

US-09-997-610-2\_COPY\_18\_45 (1-28) x AA934212 (1-458)

QY 1 AtAGlyProProLahIsProArgProProGluGluValGlyProProGlyAlaProGly 20

Db 311 GCGGACACACAGTACCCCGTAACCTGTAGACAGTAACCTGAGCTCGGC 370  
QY 21 LeuProGIntyThrglygu 27  
|||||  
Db 371 TTACCCGCAACCTGATAA 391

RESULT 4  
LOCUS  
DEFINITION  
AA934261 531 bp mRNA linear EST 26-MAR-1999  
SMOVL3CAN12B08 Onchocerca volvulus infective larva cDNA  
(SAM94WL-OVL3) Onchocerca volvulus cDNA clone onch596 5' similar to  
WP:F364.10 CE07185 CUTICULAR COLLAGEN ; contains element MSRI  
repetitive element ;, mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AA934261  
AA934261.1 GI:3091418  
EST.  
Onchocerca volvulus.  
Onchocerca volvulus.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
Onchocercidae; Onchocerca.  
1 (bases 1 to 531)  
Williams, S.A., Lizotte-Waniewski, M., Laney, S., Wenhong, L., Hillier,  
L., Allen, M., Bowles, L., Geisel, S., Jost, S., Kucaba, T., Martin, J.,  
Stephens, M., Theising, B., White, Y., Wylie, T., Chappell, J., Person, B.,  
Gibbons, M., Harvey, N., Pape, D., Chamberlain, A., Morales, R.,  
Schurk, R., Ritter, E., Kohn, S., Underwood, K. and Marra, M.  
Molecular Parasitology OVL3  
Unpublished (1998)  
Contact: Steven A. Williams  
Molecular Parasitology  
Smith College Department of Biological Sciences  
Department of Biological Sciences, Clark Science Center, Smith  
College, Northampton, MA, 01063, USA  
Tel: 4135853826  
Fax: 4135853786  
Email: genome@smith.edu  
The library was constructed by Wenhong Lu. The library is available  
from Dr. S.A. Williams, email genome@smith.edu When requesting this  
clone from Dr. Williams, please reference the Williams lab clone id  
- SMOVL3CAN12B08  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 485.  
Location/Qualifiers

FEATURES  
source  
1..531  
/organism="Onchocerca volvulus"  
/strain="Sierra Leone"  
/db\_xref="taxon:6282"  
/clone="Onch596"  
/clone\_11b="Onchocerca volvulus infective larva cDNA  
(SAM94WL-OVL3)"  
/lab\_host="XLI-Blue MRF"  
/note="Vector: lambda Unizap XR; Site\_1: Ecor I; Site\_2:  
Xho I; Cutaneous filarial nematode parasite of humans.  
mRNA was prepared from third stage infective larvae of  
Onchocerca volvulus isolated from mosquitoes 10 days after  
infection and converted to double stranded cDNA using  
reverse transcriptase and oligo(dT) followed by RNase H  
and DNAPol I. The library had 1.8 x 10<sup>5</sup> independent  
recombinants and average insert size was 900 base pairs.  
The library was constructed by Wenhong Lu. The library is  
available from Dr. S.A. Williams, email genome@smith.edu."

BASE COUNT 142 a 130 c 139 g 115 t

ORIGIN

Alignment Scores:  
Pred. No.: 46 Length: 531  
Score: 88.00 Matches: 16  
Percent Similarity: 66.67% Conservative: 2  
Best Local Similarity: 59.26% Mismatches: 9  
Query Match: 54.32% Indels: 0  
Gaps: 0



KEYWORDS	EST.
SOURCE	COW
ORGANISM	Bos taurus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
AUTHORS	1 (bases 1 to 499) Sonstegard,T.S., Capuco,A.V., Van Tassel,C.P., Ashwell,M.S. and Wells,K.D.
TITLE	Mapping of Expressed Sequence Tags from a normalized bovine mammary gland cDNA library
JOURNAL	Unpublished (2000)
COMMENT	Contact: Sonstegard TS USA, ARS, Beltsville Agricultural Research Center Bldg. 200, Rm 2A, Beltsville, MD 20705, USA Tel.: 301 504 8416 Fax: 301 504 8414 Email: tads@nrl.barc.usda.gov Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options. PCR Primers FORWARD: AGGAACAGCATGACCAT BACKWARD: GTTTCAGTCACGACG Plate: 13 row: B column: 18 Seq primer: ATTTAGGTGACACTATAC.
FEATURES	Location/Qualifiers 1..499 /organism="Bos taurus" /db_xref="taxon:9913" /clone_lib="BARC_5BOV" /tissue_type="pooled" /lab_host="DH10B" /note="Vector: PCMV SPORT6, Site_1: NotI; Site_2: SalI; Library made from pooled mRNA isolated from mammary tissues at eight physiological, developmental, and disease states."
BASE COUNT	105 a 160 c 129 g 105 t
ORIGIN	
Alignment Scores:	
Pred. No.:	65.7 Length: 499
Score:	86.00 Matches: 16
Percent Similarity:	61.54% Conservative: 0
Best Local Similarity:	61.54% Mismatches: 10
Query Match:	53.09% Indels: 0
DB:	10 Gaps: 0
US-09-997-610-2_COPY_18_45 (1-28) x BE483103 (1-499)	
OY 1	AlaglyProProlanHisProArgProProGluGluValGlyProProGlyAlaProgly 20
Db 386	GCTGGTCTCTCGTCGGCCACCCGCCCTGGCACACTGTGGGCATCTGTGGCCCTGGC 445
OY 21	LeuProGlnTyrrHngly 26
Db 446	GCTCCAGGATACCAAGT 463
RESULT 8	
LOCUS	AA847768/c 501 bp mRNA linear EST 31-MAR-1998
DEFINITION	oe91f07.f1 NCI-CGAP Col12 Homo sapiens cDNA clone IMAGE:1419781
Accession	oe91f07 to tr:063624 063624 CTD-BINDING SR-LIKE PROTEIN RAJ.; , mRNA sequence.
VERSION	AA847768
KEYWORDS	AA847768.1 GI:2934286
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 501) NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .

JOURNAL COMMENT

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cga@nci.nih.gov  
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: Stratagene, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
[www.bio.lnl.gov/bbrp/image/image.html](http://www.bio.lnl.gov/bbrp/image/image.html)  
Insert Length: 1068 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 339.

FEATURES

SOURCE

1..501  
location/Qualifiers  
organism="Homo sapiens"  
db\_xref="taxon:9606"  
clone\_image="1419781"  
clone\_id="NCI\_CGAP\_Col12"  
sex="mixed"  
tissue\_type="colon tumor"  
lab\_host="SOAR (kanamycin resistant)"  
note="Organ: colon; Vector: Bluescript SK-. Site\_1: EcoRI / Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Pooled colon tumors. 5' adaptor sequence: 5'-GAATTGGCAGCAG 3' 3' adaptor sequence: 5'-CTCCAGTTTTTTTTTTTTTTTTT 3' Average insert size: 1.2 kb." BASE COUNT 91 a 126 c 186 g 98 t ORIGIN

Alignment Scores:

Pred. No.: 65.9 Length: 501  
Score: 86.00 Matches: 16  
Percent Similarity: 73.91% Conservative: 1  
Best Local Similarity: 69.57% Mismatch: 4  
Query Match: 53.09% Indels: 2  
DB: 9 Gaps: 1

US-09-997-610-2\_COPY\_18\_45 (1-28) x AA847768 (1-501)

OY 2 GIYPPTOALNHSPTAATGPBOGUGLUAVALGIYPTOPRO-----GLYALAPRO 19  
||||||| | ||||||| :||| ||||| ||| |||  
Db 333 GGACCCCCCACGGCCGCCACGGCCCAAGAAGCAGGAGGCCCGACAGACAGGTGGCCG 274

OY 20 GIYLEUPRO 22  
||||||| |  
Db 273 GGCCCGCCC 265

RESULT 9  
BE846038 LOCUS BE846038 524 bp mRNA linear EST 25-SEP-2000  
DEFINITION 232389 BANC 5BOV Bos taurus CDNA 5', mRNA sequence.  
ACCESSION BE846038  
VERSION BE846038.1 GI:10282862  
KEYWORDS EST.  
SOURCE COW.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 524)  
Sonstegard,T.S., Capuco,A.V., Van Tassel,C.P., Ashwell,M.S. and Wells,K.D.  
Mapping of Expressed Sequence Tags from a normalized bovine mammary gland cDNA library  
Unpublished (2000)  
Contact: Sonstegard TS  
USA, ARS, Beltsville Agricultural Research Center  
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA  
Tel: 301 504 8416

Fax: 301 504 8414  
Email: tadsentr@barc.usda.gov  
Single pass sequencing. Bases called and alt-trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -mismatch 12 options.

PCR primers  
FORWARD: AGGAAACAGCTATGACCAT  
BACKWARD: GTTTCCTCCAGTCAGCAG  
Plate: 114 row: F column: 10  
Seq primer: ATTGAGTGCACCTATAG.  
Location/Qualifiers

## FEATURES

1..524  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="BARC SBOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site\_1: NotI; Site\_2: SalI;  
Library made from pooled mRNA isolated from mammary  
tissues at eight physiological, developmental, and disease  
states."  
BASE COUNT 119 a 139 c 167 g 99 t  
ORIGIN

## Alignment Scores:

Pred. No.: 68.5 Length: 524  
Score: 86.00 Matches: 16  
Percent Similarity: 61.54% Conservative: 0  
Best Local Similarity: 61.54% Mismatches: 10  
Query Match: 53.09% Indels: 0  
DB: 12 Gaps: 0

US-09-997-610-2\_COPY\_18\_45 (1-28) x BE846038 (1-524)

QY 1 AAlaglyPROProAlaHisProArgProProGluGluValGlyProProGlyAlaProGly 20  
|||||  
Db 113 GCTGTCCTCTCTGGCCCAACCGACCCCTGGCACATCTGGCCATCTGTGCTGCTG6C 172  
QY 21 LeuProGlnTyrThrGly 26  
|||  
Db 173 GCTCCAGGATTCACAAAGT 190

## RESULT 10

AVS97164 577 bp mRNA linear EST 27-NOV-2001  
AVS97164  
LOCUS AVS97164 Bos taurus cartilage fetus Bos taurus cDNA clone  
DEFINITION E1CA030C06 5', mRNA sequence.  
ACCESSION AVS97164  
VERSION AVS97164.1 GI:9714506  
KEYWORDS EST.

## SOURCE

ORGANISM Bos taurus  
COW.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 577)  
AUTHORS Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.  
and Sugimoto,Y.  
TITLE Establishment of a high throughput EST sequencing system using  
poly(A) tail-removed cDNA libraries and determination of 36,000  
bovine ESTs

JOURNAL Nucleic Acids Res. 29 (22), E108 (2001)  
MEDLINE 21570554  
COMMENT Contact: Yoshikazu Sugimoto  
Animal Genetics Division  
Shirakawa Institute of Animal Genetics  
OdaKura, Nishigo, Nishi-Shirakawa, Fukushima 961-8061, Japan  
Tel: 81-248-25-5641  
Fax: 81-248-25-5725  
Email: kazusugie@ccoc.ocn.ne.jp  
Single pass sequencing.  
This clone was obtained from a polyA-deleted cDNA library.  
Location/Qualifiers

## FEATURES

## source

1..577  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="E1CA030C06"  
/clone\_lib="Bos taurus cartilage fetus"  
/tissue\_type="cartilage"  
/dev\_stage="fetus"  
/lab\_host="DH10B"  
/note="Vector: pZL1; Site\_1: SalI; Site\_2: NotI; Poly A  
was deleted from a NotI site"  
BASE COUNT 123 a 170 c 171 g 113 t  
ORIGIN

## Alignment Scores:

Pred. No.: 74.5 Length: 577  
Score: 86.00 Matches: 16  
Percent Similarity: 61.54% Conservative: 0  
Best Local Similarity: 61.54% Mismatches: 10  
Query Match: 53.09% Indels: 0  
DB: 10 Gaps: 0

US-09-997-610-2\_COPY\_18\_45 (1-28) x AVS97164 (1-577)

QY 1 AAlaglyPROProAlaHisProArgProProGluGluValGlyProProGlyAlaProGly 20  
|||||  
Db 245 GCTGTCCTCTCTGGCCCAACCGACCCCTGGCACATCTGGCCATCTGTGCTGCTG6C 304  
QY 21 LeuProGlnTyrThrGly 26  
|||  
Db 305 GCTCCAGGATTCACAAAGT 322

## RESULT 11

AW257669 662 bp mRNA linear EST 20-DEC-1999  
AW257669  
LOCUS SWYD25CAU08D04SK Brugia malayi young adult day 25 cDNA  
DEFINITION (SAM99MLW-BMYD25) Brugia malayi cDNA clone SWYD25CAU08D04 5', mRNA  
sequence.  
ACCESSION AW257669  
VERSION AW257669.1 GI:6605842  
KEYWORDS EST.

## SOURCE

ORGANISM Brugia malayi.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidae;  
Onchocercidae; Brugia.  
REFERENCE 1 (bases 1 to 662)  
AUTHORS Williams,S.A.  
TITLE Genes expressed in young adult day 25 of Brugia malayi  
JOURNAL Unpublished (1999)  
COMMENT Contact: Steven A. Williams  
Molecular Parasitology  
Smith College Department of Biological Sciences  
Department of Biological Sciences, Clark Science Center, Smith  
College, Northampton, MA, 01063, USA  
Tel: 4135853826  
Fax: 4135853786  
Email: genome@smith.edu  
Seq primer: pBluescript SK.

## FEATURES

1..662  
Location/Qualifiers  
/organism="Brugia malayi"  
/db\_xref="taxon:6279"  
/clone\_lib="SWYD25CAU08D04"  
/clone\_lib="Brugia malayi young adult day 25 cDNA  
(SAM99MLW-BMYD25)"  
/dev\_stage="young adult, twenty five days after infection"  
/lab\_host="XLI-Blue MRF"  
/note="Vector: Lambda Uni-ZAP XR; Site\_1: Eco RI; Site\_2:  
Xho I; Lymphatic filarial nematode parasite of humans.  
mRNA was prepared from young adult worms isolated from  
the peritoneal cavity of jirds on day 25 after infection  
and converted to double-stranded cDNA using reverse  
transcriptase and oligo(dt) followed by RNase H and DNA  
pol I. The library has 6.2 x 10<sup>5</sup> independent recombinants

and the average insert size is approx. 110bp. The library was constructed by Michelle Lizotte-Waniewski. The library is available from Dr. S.A. Williams, email: genome@neal.smith.edu."

BASE COUNT 195 a 176 c 165 g 120 t 6 others

ORIGIN

Alignment Scores:

Pred. No.:	83.8	Length:	662
Score:	86.00	Matches:	18
Percent Similarity:	54.55%	Conservative:	0
Best Local Similarity:	54.55%	Mismatches:	9
Query Match:	53.09%	Indels:	6
DB:	10	Gaps:	1

US-09-997-610-2\_COPY\_18\_45 (1-28) x AW257669 (1-662)

QY 1 AAlaGlyProProAlaHisProArg-----ProProGluGluValGly 14

DB 458 GCCGACACACCTGGCAGACCAGCAATAGTGCATGCTGTCACACGAGCCGTTGGGA 517

QY 15 PROProGlyAlaProGlyLeuProGlnTyrThrGlyGlu 27

DB 518 CCACGAGAGACACGAGTCAGCCAGGAGCTGACGCGGAA 556

RESULT 12 609 bp mRNA linear EST 27-APR-2000

LOCUS AW675980 SWD25CAU013F09SK Brugia malayi young adult day 25 CDNA

DEFINITION (SAM99MLW-BmyD25) Brugia malayi cDNA clone SWD25CAU013F09 5', mRNA sequence.

ACCESSION AW675980

VERSION AW675980

KEYWORDS EST

SOURCE Brugia malayi.

ORGANISM Brugia malayi.

REFERENCE 1 (bases 1 to 609)

AUTHORS Williams, S.A.

TITLE Genes expressed in young adult day 25 of Brugia malayi

JOURNAL Unpublished (1999)

COMMENT Contact: Steven A. Williams

Smith College Department of Biological Sciences

Department of Biological Sciences, Clark Science Center, Smith

College, Northampton, MA, 01063, USA

Tel: 4135853826

Fax: 4135853786

Email: genome@smith.edu

Seq primer: pbluescript SK.

Location/Qualifiers

1. 609

/organism="Brugia malayi"

/db\_xref="taxon:6279"

/clone="SWD25CAU013F09"

/clone\_1lb="Brugia malayi young adult day 25 CDNA

(SAM99MLW-BmyD25)"

/dev\_stage="young adult, twenty five days after infection"

/lab\_host="XLI-Blue MRF"

/note="Vector: Lambda Uni-ZAP XR; Site.1: Eco RI; Site.2:

Xho I; Lymphatic filarial nematode parasite of humans.

mRNA was prepared from young adult worms isolated from

the peritoneal cavity of jirds on day 25 after infection

and converted to double-stranded cDNA using reverse

transcriptase and oligo(dT) followed by RNase H and DNA

pol I. The library has 6.2 x 10<sup>5</sup> independent recombinants

and the average insert size is approx. 110bp. The library

was constructed by Michelle Lizotte-Waniewski. The

library is available from Dr. S.A. Williams, email:

genome@neal.smith.edu."

genome@neal.smith.edu."

179 a 162 c 155 g 113 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.:	86.4	Length:	609
Score:	85.50	Matches:	18
Percent Similarity:	56.25%	Conservative:	0
Best Local Similarity:	56.25%	Mismatches:	9
Query Match:	52.78%	Indels:	5
DB:	10	Gaps:	1

US-09-997-610-2\_COPY\_18\_45 (1-28) x AW675980 (1-609)

QY 1 AAlaGlyProProAlaHisPro-----ArgProProGluGluValGlyPro 15

DB 439 GCCGACACACCTGGCAGACCAGCAATAGCATTGCTGTCACACGAGCCGTTGGACA 498

QY 16 PROGlyAlaProGlyLeuProGlnTyrThrGlyGlu 27

DB 499 CCAGGAGACACGAGTCAGCCAGGAGCTGACGCGGAA 534

RESULT 13

LOCUS AW347951 626 bp mRNA linear EST 01-FEB-2000

DEFINITION (SAM99MLW-BmyD25) Brugia malayi cDNA clone SWD25CAU09B11 5', mRNA sequence.

ACCESSION AW347951

VERSION AW347951

KEYWORDS EST

SOURCE Brugia malayi.

ORGANISM Brugia malayi.

REFERENCE 1 (bases 1 to 626)

AUTHORS Williams, S.A.

TITLE Genes expressed in young adult day 25 of Brugia malayi

JOURNAL Unpublished (1999)

COMMENT Contact: Steven A. Williams

Molecular Parasitology

Smith College Department of Biological Sciences

Department of Biological Sciences, Clark Science Center, Smith

College, Northampton, MA, 01063, USA

Tel: 4135853826

Fax: 4135853786

Email: genome@smith.edu

Seq primer: pbluescript SK.

Location/Qualifiers

1. 626

/organism="Brugia malayi"

/db\_xref="taxon:6279"

/clone="SWD25CAU09B11"

/clone\_1lb="Brugia malayi young adult day 25 CDNA

(SAM99MLW-BmyD25)"

/dev\_stage="young adult, twenty five days after infection"

/lab\_host="XLI-Blue MRF"

/note="Vector: Lambda Uni-ZAP XR; Site.1: Eco RI; Site.2:

Xho I; Lymphatic filarial nematode parasite of humans.

mRNA was prepared from young adult worms isolated from

the peritoneal cavity of jirds on day 25 after infection

and converted to double-stranded cDNA using reverse

transcriptase and oligo(dT) followed by RNase H and DNA

pol I. The library has 6.2 x 10<sup>5</sup> independent recombinants

and the average insert size is approx. 110bp. The library

was constructed by Michelle Lizotte-Waniewski. The

library is available from Dr. S.A. Williams, email:

genome@neal.smith.edu."

genome@neal.smith.edu."

189 a 164 c 159 g 114 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.:	88.5	Length:	626
Score:	85.50	Matches:	18
Percent Similarity:	56.25%	Conservative:	0
Best Local Similarity:	56.25%	Mismatches:	9

Query Match: 52.78% Indels: 5  
DB: 10 Gaps: 1  
US-09-997-610-2\_COPY\_18\_45 (1-28) x AW347951 (1-626)  
QY 1 ALaGlyProProAlaHsPro-----ArgProProGluGluValGlyPro 15  
|||||  
Db 459 GCCGACCCACTGGCAGACCAAGAAATAGCAATGCTGTGTCCACAGAGACCGCTGGGACCA 518  
QY 16 ProGlyAlaProGlyLeuProGlnTyrThrGlyGlu 27  
|||||  
Db 519 CCAGGAGCACCGTGCACCGACGAGAGCTGACGGCGAA 554  
RESULT 14  
H30955 391 bp mRNA linear EST 12-DEC-1996  
LOCUS SW31CA178SK Brugia malayi infective larva cDNA (SAM94WL-Bml3)  
DEFINITION Brugia malayi cDNA clone SW31CA178 5' similar to cuticular collagen  
1, mRNA sequence.  
ACCESSION H30955  
VERSION H30955.1 GI:902111  
KEYWORDS EST.  
SOURCE Brugia malayi.  
ORGANISM Brugia malayi  
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
Onchocercidae; Brugia.  
REFERENCE 1 (bases 1 to 391)  
AUTHORS Blaxter, M.L., Raghavan, N., Ghosh, I., Gulliano, D., Lu, W., Williams, S.A., Slatko, B. and Scott, A.L.  
TITLE Genes expressed in Brugia malayi infective third stage larvae  
JOURNAL Mol. Biochem. Parasitol. 77 (1), 77-93 (1996)  
MEDLINE 96379223  
COMMENT Contact: Steven A. Williams  
Molecular Parasitology  
Smith College Department of Biological Sciences  
Department of Biological Sciences, Clark Science Center, Smith  
College, Northampton, MA, 01063, USA  
Tel: 4135853826  
Fax: 4135853786  
Email: genome@smith.edu  
Seq primer: pBluescript SK.  
FEATURES  
source  
1..391  
/organism="Brugia malayi"  
/strain="TRS Labs"  
/db\_xref="taxon:6279"  
/clone="SW31CA178"  
/clone\_lib="Brugia malayi infective larva cDNA  
(SAM94WL-Bml3)"  
/lab\_host="XLI-Blue MRP"  
/note="vector: lambda Unizap XR; Site.1: EcoR I; Site.2:  
Xho I; Lymphatic filarial nematode parasite of humans  
mRNA was prepared from third stage infective larvae of  
Brugia malayi isolated from mosquitoes 10 days after  
infection and converted to double stranded cDNA using  
reverse transcriptase and oligo(dT) followed by RNase H  
and DNase I. The library had 1.6 x 10<sup>6</sup> independent  
recombinants and average insert size was 900 base pairs.  
The library was constructed by Wenhong Lu. The library is  
available from Dr. S.A. Williams, email genome@smith.edu."  
BASE COUNT 100 a 106 c 119 g 60 t  
ORIGIN  
Alignment Scores: 65.4 Length: 391  
Pred. No.: 18  
Score: 85.00 Matches: 18  
Percent Similarity: 54.55% Conservative: 0  
Best Local Similarity: 54.55% Mismatches: 9  
Query Match: 52.47% Indels: 6  
DB: 14 Gaps: 1  
US-09-997-610-2\_COPY\_18\_45 (1-28) x H30955 (1-391)

QY 1 ALaGlyProProAlaHsPro-----ArgProProGluGluValGly 14  
|||||  
Db 35 GCCGACCCACTGGCAGACCAAGAAATAGTGTATGCTGTGTCCACAGAGCCGCTGGGA 94  
QY 15 ProProGlyAlaProGlyLeuProGlnTyrThrGlyGlu 27  
|||||  
Db 95 CCACGAGAGCACCGTGCACCGACGAGAGCTGACGGCGAA 133  
RESULT 15  
AA585547 490 bp mRNA linear EST 30-DEC-1997  
LOCUS SW3D9CA131SK Brugia malayi L3 molting-day 9 larva cDNA  
DEFINITION (SAM97MLW-Bml3d9) Brugia malayi cDNA clone SW3D9CA131 5', mRNA  
sequence.  
ACCESSION AA585547  
VERSION AA585547.1 GI:2392959  
KEYWORDS EST.  
SOURCE Brugia malayi.  
ORGANISM Brugia malayi  
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
Onchocercidae; Brugia.  
REFERENCE 1 (bases 1 to 490)  
AUTHORS Williams, S.A. and Lizotte-Waniewski, M.  
TITLE Genes expressed in day 9 post-infection, third stage larvae of  
Brugia malayi  
JOURNAL Unpublished (1997)  
COMMENT Contact: Steven A. Williams  
Molecular Parasitology  
Smith College Department of Biological Sciences  
Department of Biological Sciences, Clark Science Center, Smith  
College, Northampton, MA, 01063, USA  
Tel: 4135853826  
Fax: 4135853786  
Email: genome@smith.edu  
Seq primer: pBluescript SK.  
FEATURES  
source  
1..490  
/organism="Brugia malayi"  
/db\_xref="taxon:6279"  
/clone="SW3D9CA131"  
/clone\_lib="Brugia malayi L3 molting-day 9 larva cDNA  
(SAM97MLW-Bml3d9)"  
/dev\_stage="third stage larvae, nine days after infection"  
/lab\_host="E. coli XLI-Blue MRP"  
/note="vector: LambdaZap II (Unizap XR); Site.1: Eco RI  
(5' end); Site.2: Xho I (3' end); Brugia malayi is a  
lymphatic filarial nematode parasite of humans. mRNA was  
prepared from third stage larvae of Brugia malayi  
isolated from the peritoneal cavity of jirds nine days  
after infection. The mRNA was converted to double  
stranded cDNA using reverse transcriptase and oligo (dT)  
followed by RNase H and DNase I. The library was  
constructed by Michelle Lizotte-Waniewski. The library is  
available from the Filarial Genome Project Resource  
Center: contact Dr. S.A. Williams, Clark Science Center,  
Smith College, Northampton, MA 01063 USA phone +1 413  
585-3826 fax +1 413 585-3786 email genome@smith.edu"  
BASE COUNT 135 a 124 c 125 g 105 t  
ORIGIN  
Alignment Scores: 79.4 Length: 490  
Pred. No.: 18  
Score: 85.00 Matches: 18  
Percent Similarity: 54.55% Conservative: 0  
Best Local Similarity: 54.55% Mismatches: 9  
Query Match: 52.47% Indels: 6  
DB: 9 Gaps: 1  
US-09-997-610-2\_COPY\_18\_45 (1-28) x AA585547 (1-490)  
QY 1 ALaGlyProProAlaHsPro-----ArgProProGluGluValGly 14  
|||||  
Db 6 GCCGACCCACTGGCAGACCAAGAAATAGTGTGTGTCCACGAGACCGCTGGGA 65

Qy 15 ProProGlyAlaProGlyLeuProGlnTyrThrGlyGlu 27  
|||||  
Db 66 CCACCAGGAGCACACAGGTCAGCCAGGAGCTGACGGCGAA 104

Search completed: February 19, 2003, 19:36:01  
Job time : 219.528 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 19, 2003, 14:57:18 ; Search time 390.136 Seconds  
(without alignments)  
2088.706 Million cell updates/sec

Title: US-09-997-610-2\_COPY\_18\_45  
Perfect score: 162  
Sequence: 1 AGPPAHPRPPEVGPFGAPGLPYGTGEI 28

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -DEV\_xlh  
-MODEL=frame+p2n.model -DEV\_xlh  
-O/cgn2.1/USPTO/spool/US09997610/runat\_10022003\_160823\_23819/app\_query.fasta.1.1635  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human4.0.cdi -LIST=45  
-DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=fto -NORM-ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09997610\_@cgn.1.1.3965-runat\_10022003\_160823\_23819 -NCPU=6 -ICPU=3  
-NO\_MLPTX -NO\_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl :  
1: gb\_da : \*  
2: gb\_hcg : \*  
3: gb\_in : \*  
4: gb\_om : \*  
5: gb\_ov : \*  
6: gb\_pat : \*  
7: gb\_ph : \*  
8: gb\_pl : \*  
9: gb\_pr : \*  
10: gb\_ro : \*  
11: gb\_sts : \*  
12: gb\_sy : \*  
13: gb\_un : \*  
14: gb\_vl : \*  
15: em\_da : \*  
16: em\_fun : \*  
17: em\_hum : \*  
18: em\_in : \*  
19: em\_mu : \*  
20: em\_om : \*  
21: em\_or : \*  
22: em\_ov : \*  
23: em\_pat : \*  
24: em\_ph : \*  
25: em\_pl : \*  
26: em\_ro : \*  
27: em\_sts : \*  
28: em\_un : \*

29: em\_vl : \*  
30: em\_hcg\_hum : \*  
31: em\_hcg\_inv : \*  
32: em\_hcg\_other : \*  
33: em\_hcg\_mus : \*  
34: em\_hcg\_pin : \*  
35: em\_hcg\_rtd : \*  
36: em\_hcg\_mam : \*  
37: em\_hcg\_vrt : \*  
38: em\_sy : \*  
39: em\_hcgo\_hum : \*  
40: em\_hcgo\_mus : \*  
41: em\_hcgo\_other : \*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	73.5	145880	9 HS30209	Z82198 Human DNA s
2	86	53.1	4428	6 AX146422	AX146422 Sequence
3	86	53.1	4428	6 AX146424	AX146424 Sequence
4	84.5	52.2	68661	3 AC024805	AC024805 Caenorhab
5	84.5	52.2	278007	2 AC006799	AC006799 Caenorhab
6	84	51.9	4428	6 AX146430	AX146430 Sequence
7	84	51.9	66669	3 AF017777	AF017777 Drosophill
8	84	51.9	87089	2 AC013980	AC013980 Drosophill
9	84	51.9	171548	3 AC092494	AC092494 Drosophill
10	84	51.9	173508	3 AC011251	AC011251 Drosophill
11	84	51.9	256417	2 AC079422	AC079422 Mus muscu
12	84	51.9	292911	3 AE003568	AE003568 Drosophill
13	83	51.2	186	6 AR014097	AR014097 Sequence
14	83	51.2	186	6 AR117066	AR117066 Sequence
15	83	51.2	186	6 I06482	I06482 Sequence 19
16	83	51.2	186	6 I20524	I20524 Sequence 62
17	83	51.2	287	3 O0S420010	AJ420010 Osterlag1
18	83	51.2	3612	10 MMCOLA4	X067777 Mouse mRNa
19	83	51.2	6512	6 AX306181	AX306181 Sequence
20	83	51.2	6512	10 MUSCOL1A4A	J04694 M. musculus
21	82.5	50.9	437	10 MMA379COL	X91012 M. musculus
22	82.5	50.9	1009	9 BC004412	BC004412 Homo sapi
23	82.5	50.9	1521	10 AF237721	AF237721 Mus muscu
24	82.5	50.9	1605	10 BC030945	BC030945 Mus muscu
25	82.5	50.9	2091	10 BC007574	BC007574 Homo sapl
26	82.5	50.9	2388	10 AF349718	AF349718 Mus muscu
27	82.5	50.9	2437	9 BC011705	BC011705 Homo sapl
28	82.5	50.9	2480	9 HDWCOL3IX	L41162 Homo sapien
29	82.5	50.9	2543	6 AR014074	AR014074 Sequence
30	82.5	50.9	6246	6 AR111836	AR111836 Sequence
31	82.5	50.9	6246	6 HSCOL9A3S2	AF026802 Homo sapl
32	82.5	50.9	69252	2 AC101519	AC101519 Mus muscu
33	82.5	50.9	160241	9 HS88517	AL035669 Human DNA
34	82.5	50.9	226854	10 AL669926	AL669926 Mouse DNA
35	82	50.6	426	9 AF230926	AF230926 Macaca mu
36	82	50.6	964	6 AX067354	AX067354 Sequence
37	82	50.6	1485	10 MUSC3A1A	M18933 Mouse alpha
38	82	50.6	1881	6 A60690	A60690 Sequence 20
39	82	50.6	1881	6 AR123971	AR123971 Sequence
40	82	50.6	3234	9 HSC3A1R	X15332 Human COL3A
41	82	50.6	3902	9 BC028178	BC028178 Homo sapl
42	82	50.6	5460	6 AX302553	AX302553 Sequence
43	82	50.6	5460	6 AX393295	AX393295 Sequence
44	82	50.6	5460	6 E10600	E10600 cDNA encod1
45	82	50.6	5460	9 HSCOL3A1	X14420 Human mRNa

## ALIGNMENTS

RESULT 1

HS302D9  
LOCUS HS302D9 145880 bp DNA linear PRI 12-DEC-1999  
DEFINITION Human DNA sequence from clone RP1-302D9 on chromosome 22 Contains  
GSSS, complete sequence.  
ACCESSION Z82198  
VERSION Z82198  
KEYWORDS HNC  
SOURCE HG:6572207  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 145880)  
AUTHORS Bridgeman, A.  
TITLE Direct Submission  
JOURNAL Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Dec 13, 1999 this sequence version replaced gi:3164067.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em:, EMBL; SW:, SWISSPROT; Tr:, TREMBL; Wp:, WormPEP; Information  
on the WormPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
Chromosome 22, constructed by the Sanger Centre Chromosome 22  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr22  
RP1-302D9 is from the library RPCT-1 constructed at the Roswell  
Park Cancer Institute by the group of Pieter de Jong. For further  
details see http://bacpac.med.buffalo.edu/  
VECTOR: pCYPAC2  
This sequence is the entire insert of clone RP1-302D9 The true left  
end of clone CTA-282F2 is at 69682 in this sequence. The true right  
end of clone CTA-415G2 is at 55167 in this sequence.  
FEATURES  
source  
1..145880  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="22"  
/clone="RP1-302D9"  
/clone\_lib="RPCT-1"  
188..245  
/note="MER3 repeat: matches 144..209 of consensus"  
246..571  
/note="AluX repeat: matches 1..312 of consensus"  
572..759  
/note="MER3 repeat: matches 1..144 of consensus"  
783..933  
/note="MER5A repeat: matches 26..187 of consensus"  
1033..1336  
/note="AluSP repeat: matches 1..299 of consensus"  
1450..1583  
/note="MIR repeat: matches 24..160 of consensus"  
1687..1752  
/note="L2 repeat: matches 2593..2661 of consensus"  
2350..2660  
/note="AluSC repeat: matches 3..309 of consensus"  
2684..2981  
/note="AluSq repeat: matches 2..300 of consensus"  
3323..3343  
/note="MUTIE repeat: matches 116..136 of consensus"

repeat\_region 3344..3652  
/note="AluY repeat: matches 1..309 of consensus"  
repeat\_region 3653..3928  
/note="MUTIE repeat: matches 136..359 of consensus"  
repeat\_region 3929..4278  
/note="THEB repeat: matches 3..364 of consensus"  
4279..4485  
/note="MUTIE repeat: matches 359..568 of consensus"  
5073..5176  
/note="52 copies 2 mer ct 78 conserved"  
5181..5491  
/note="AluB repeat: matches 1..311 of consensus"  
6369..6485  
/note="L2 repeat: matches 2579..2705 of consensus"  
6647..6685  
/note="MAD1 repeat: matches 1..23 of consensus"  
6686..6987  
/note="AluX repeat: matches 1..302 of consensus"  
6988..7036  
/note="MAD1 repeat: matches 23..77 of consensus"  
7482..7754  
/note="AluB repeat: matches 9..290 of consensus"  
7775..8060  
/note="AluB repeat: matches 1..295 of consensus"  
8414..8551  
/note="L2 repeat: matches 2553..2706 of consensus"  
8914..9030  
/note="MIR repeat: matches 147..262 of consensus"  
9110..9280  
/note="MIR repeat: matches 91..262 of consensus"  
9283..9412  
/note="MIR repeat: matches 15..144 of consensus"  
9521..9679  
/note="FAM repeat: matches 3..161 of consensus"  
9820..10225  
/note="MSTB repeat: matches 2..425 of consensus"  
complement(10179..10678)  
/note="match: GSS: Em: B56592"  
complement(10204..10728)  
/note="match: GSS: Em: A0701486"  
complement(10249..10706)  
/note="match: GSS: Em: A0225455"  
10312..10383  
/note="MIR repeat: matches 79..150 of consensus"  
10718..11310  
/note="match: GSS: Em: B14024"  
10784..11201  
/note="match: GSS: Em: B43656"  
11858..11946  
/note="MIR repeat: matches 20..137 of consensus"  
12174..12445  
/note="L2 repeat: matches 1988..2275 of consensus"  
12444..12642  
/note="MIR repeat: matches 63..241 of consensus"  
13017..13369  
/note="match: STR: Em: G49301"  
13331..13397  
/note="MIR repeat: matches 174..244 of consensus"  
13398..13698  
/note="AluSP repeat: matches 1..302 of consensus"  
13699..13810  
/note="MIR repeat: matches 76..174 of consensus"  
13806..13919  
/note="MIR repeat: matches 77..189 of consensus"  
13945..14060  
/note="MIR repeat: matches 24..142 of consensus"  
14061..14367  
/note="AluY repeat: matches 1..301 of consensus"  
14368..14452  
/note="MIR repeat: matches 141..225 of consensus"  
14589..14679  
/note="MIR repeat: matches 173..262 of consensus"  
14597..15201  
misc\_feature

misc_feature	/note="match: GSS: Em: A0553482"	14616. .15060	
repeat_region	/note="match: GSS: Em: A0370601"	14868. .15040	
repeat_region	/note="MIR repeat: matches 49. .233 of consensus"	15071. .15188	
repeat_region	/note="L2 repeat: matches 2112. .2239 of consensus"	15304. .15399	
repeat_region	/note="MLT1B repeat: matches 1. .99 of consensus"	15490. .15662	
repeat_region	/note="AluSg1 repeat: matches 2. .114 of consensus"	15669. .15727	
repeat_region	/note="MLT1B repeat: matches 119. .178 of consensus"	15788. .16027	
repeat_region	/note="AluSc repeat: matches 1. .299 of consensus"	16028. .16245	
repeat_region	/note="MLT1B repeat: matches 178. .390 of consensus"	16566. .16854	
repeat_region	/note="AluY repeat: matches 1. .300 of consensus"	18296. .18323	
repeat_region	/note="MSTR repeat: matches 2. .29 of consensus"	18324. .18392	
repeat_region	/note="MER66-internal repeat: matches 4919. .4993 of consensus"	18393. .18712	
repeat_region	/note="Aluub repeat: matches 1. .311 of consensus"	18713. .19133	
repeat_region	/note="MER66-internal repeat: matches 4548. .4919 of consensus"	complement(18872. .19230)	
misc_feature	/note="match: GSS: Em: A0005063"	19251. .19719	
misc_feature	/note="match: GSS: Em: B14179"	19537. .20290	
repeat_region	/note="HERVFN21 repeat: matches 4657. .5784 of consensus"	20317. .20382	
repeat_region	/note="33 copies 2 mer ta 68 conserved"	20513. .20666	
repeat_region	/note="77 copies 2 mer tt 70 conserved"	20682. .21008	
repeat_region	/note="AluSg1 repeat: matches 1. .306 of consensus"	21239. .21553	
repeat_region	/note="HUES-P3 repeat: matches 4410. .4713 of consensus"	21882. .22254	
repeat_region	/note="THE1B repeat: matches 1. .364 of consensus"	22302. .22537	
repeat_region	/note="MER66-internal repeat: matches 2186. .2417 of consensus"	22538. .22850	
repeat_region	/note="AluSp repeat: matches 1. .313 of consensus"	22851. .23801	
repeat_region	/note="MER66-internal repeat: matches 1210. .2186 of consensus"	23905. .23989	
repeat_region	/note="MER66-internal repeat: matches 3017. .3102 of consensus"		
Alignment Scores:			
Pred. NO.:	2.37	Length:	145880
Score:	119.00	Matches:	20
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	73.46%	Indels:	0
DB:	9	Gaps:	0
US-09-997-610-2_COPY_18_45 (1-28) x HS30209 (1-145880)			
OY	1	AlaGlyProGdaLahNsProArGrProProGluGluValAlGlyProProGluAlaProGly	20
Db	34837	GCTGACACCTCCAGCACCACCCAGGCCCCAGAAAGATGGGCGCTCTGTGCACACAGT	34896
RESULT 2			
AX146422			
LOCUS	AX146422	4428 bp	DNA linear PAT 31-MAY-2001

DEFINITION	Sequence 3 from Patent WO0134647.
ACCESSION	AX146422
VERSION	AX146422.1 GI:14284845
KEYWORDS	.
SOURCE	COW.
ORGANISM	Bos taurus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
TITLE	Bovidae; Bovinae; Bos.
JOURNAL	1 (bases 1 to 4428)
	Bell,M.P., Neff,T.B., Polarek,J.W. and Seeley,T.W.
	Animal collagens and gelatins
	Patent: WO 0134647-A 3 17-MAY-2001;
	FIBROGEN, INC. (US)
FEATURES	Location/Qualifiers
source	1..4428
	/organism="Bos taurus"
	/db_xref="taxon:9913"
BASE COUNT	973 a 1185 c 1400 g 870 t
ORIGIN	
Alignment Scores:	
Pred. No.:	89.9 Length: 4428
Score:	86.00 Matches: 16
Percent Similarity:	61.54% Conservative: 0
Best Local Similarity:	61.54% Mismatches: 10
Query Match:	53.09% Indels: 0
DB:	Gaps: 0
. US-09-997-610-2_COPY_18_45 (1-28) x AX146422 (1-4428)	
OY	1 AAlAgIyPrOrFoLdAnIsPrOArPrOrFoGluGlValIgLyPrOrFoGIyAlAPrOcLy 20
Dn	538 GCTGTCTCCTCGGGCCACCAGCCCTGCACATCTGGCGCAATCTGGTCCCCTGGC 597
OY	21 LeuPrOGInTyGrThrgly 26
Dn	598 GCCTCAGATACCAAGGT 615
RESULT 3	
LOCUS	AX146424 4428 bp DNA linear PAT 31-MAY-2001
AX146424	
DEFINITION	Sequence 5 from Patent WO0134647.
ACCESSION	AX146424
VERSION	AX146424.1 GI:14284846
KEYWORDS	.
SOURCE	COW.
ORGANISM	Bos taurus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
	Bovidae; Bovinae; Bos.
REFERENCE	1 (bases 1 to 4428)
AUTHORS	Bell,M.P., Neff,T.B., Polarek,J.W. and Seeley,T.W.
TITLE	Animal collagens and gelatins
JOURNAL	Patent: WO 0134647-A 5 17-MAY-2001;
	FIBROGEN, INC. (US)
FEATURES	Location/Qualifiers
source	1..4428
	/organism="Bos taurus"
	/db_xref="taxon:9913"
BASE COUNT	974 a 1185 c 1399 g 870 t
ORIGIN	
Alignment Scores:	
Pred. No.:	89.9 Length: 4428
Score:	86.00 Matches: 16
Percent Similarity:	61.54% Conservative: 0
Best Local Similarity:	61.54% Mismatches: 10
Query Match:	53.09% Indels: 0
DB:	Gaps: 0
US-09-997-610-2_COPY_18_45 (1-28) x AX146424 (1-4428)	

OY	1	Ala1y1pro1a1h1s1p1r1o1p1r1o1g1u1s1u1a1g1y1p1r1o1g1y1a1l1a1p1r1o1g1y1	20
Db	538	gctggtctctctctgccccaccgaccccttgscacactgctgscatctctggtgccccctggc	597
OY	21	Leu1p1r1o1s1i1n1t1y1r1h1r1g1y1	26
Db	598	gcttcacagatattccagagt	615
RESULT 4			
AC024805/c			
LOCUS	AC024805	68661 bp	DNA linear INV 29-MAY-2002
DEFINITION	Caenorhabditis elegans cosmid Y51H7C, complete sequence.		
ACCESSION	AC024805		
VERSION	AC024805.1	GI:7140366	
KEYWORDS	HTG.		
SOURCE	Caenorhabditis elegans.		
ORGANISM	Caenorhabditis elegans		
	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;		
	Rhabditoidae; Rhabditidae; Pelodierinae; Caenorhabditis.		
REFERENCE	1 (bases 1 to 68661)		
AUTHORS	Waterston,R.		
TITLE	Genome sequence of the nematode C. elegans: a platform for		
JOURNAL	investigating biology. The C. elegans Sequencing Consortium		
MEDLINE	Science 282 (5396), 2012-2018 (1998)		
PUBMED	99069613		
REFERENCE	2 (bases 1 to 68661)		
AUTHORS	Bradshaw-Cordum,H. and Dubuque,T.		
TITLE	The sequence of C. elegans cosmid Y51H7C		
JOURNAL	Unpublished (2001)		
REFERENCE	3 (bases 1 to 68661)		
AUTHORS	Waterston,R.H.		
TITLE	Direct Submission		
JOURNAL	Submitted (01-MAR-2000) Genome Sequencing Center, Washington		
	University School of Medicine, 4444 Forest Park Parkway, St. Louis,		
	MO 63100, USA		
REFERENCE	4 (bases 1 to 68661)		
AUTHORS	Waterston,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (24-MAR-2000) Department of Genetics, Washington		
	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA		
REFERENCE	5 (bases 1 to 68661)		
AUTHORS	Waterston,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (24-APR-2001) Department of Genetics, Washington		
	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA		
REFERENCE	6 (bases 1 to 68661)		
AUTHORS	Waterston,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (22-AUG-2001) Department of Genetics, Washington		
	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.		
	Louis, MO 63110, USA		
REFERENCE	7 (bases 1 to 68661)		
AUTHORS	Waterston,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (07-SEP-2001) Department of Genetics, Washington		
	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.		
	Louis, MO 63110, USA		
REFERENCE	8 (bases 1 to 68661)		
AUTHORS	Waterston,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-MAY-2002) Department of Genetics, Washington		
	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.		
	Louis, MO 63110, USA		
COMMENT	Submitted by:		

It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m3 subclone.

For a graphical representation of this cosmid sequence and its analysis see:  
<http://www.wormbase.org/db/seq/sequence?name=Y51H7C.class=Sequence>

## NEIGHBORING COSMID INFORMATION

The 5' cosmid is F54D12, 200 bp overlap; the 3' cosmid is H17B01 200 bp overlap.

## NOTES

coding sequences below are the result of integration and manual review of the following data : computer analysis using the program GeneFinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yuiji Kohara ([http://www.ddbj.nig.ac.jp/c-elegans/hml/CE\\_INDEX.html](http://www.ddbj.nig.ac.jp/c-elegans/hml/CE_INDEX.html)) and The C elegans ORFome cloning project (<http://wormfb.dcl.harvard.edu/>), similarity to other proteins from Blastx analyses (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans Genbank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Dddy, S.R., 1997, Nucleic Acids. Res., 25, 955-964).

## FEATURES

## Source

```

/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db.xref="taxon:6239"
/chromosome="II"
/clone="Y51H7C"
585..862
/gene="Y51H7C.8"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence/name=Y51H7C.8; class
s=Sequence"
join(585..630,690..862)

```

```

/product="Hypochemical protein y1h7c.8"
/protein_id="AAK39332.1"
/db_xref="GI:13775524"
/translation="MSLVQYDELGLGCDGIQRKLNITVLCSIGKMEKGLRRRYV
VQSVLEKVIKRYENGCGHEQMDDEQGMELKF"
1328. .1834
gene

```

<http://www.wormbase.org/db/seq/sequence?name=Y51H7C.7;class=S=Sequence>  
join(1328, .1450,1505, .1630,1682, .1834)

```
/note= coded for by the following C. elegans cDNAs:
yk560h11.5, yk118a3.3, yk118a3.5, yk560h11.3"
/codon_start=1
```

```

/product="Hyprothetical protein Y51HC.7"
/protein_id="AAK39333.1"
/db_xref="GI:13775525"
/translation="MSRSKSTPADVIAKLQKPEDLITKLEDAVEEDVGVSEKHERRSR
HEBALVDINSPKLMDSLMAVQATKGGCAKNDLLIDLARTKRMADMKELNLRQDA
PRINKAQQTSPRLNMLQPEQGESESKAK"

```

NOTICE: This sequence may not be the entire insert of this clone.

## CDS

/note="for a graphical representation of this gene see:  
http://www.wormbase.org/db/seq/sequence?name=Y51H7C.6a;ccla  
ss=Sequence"  
join(2368..2797,2854..2979,3032..3261,4332..4496,  
4570..5048,5864..6340,6802..7129,7558..7728)  
/gene="Y51H7C.6"

/note="coded for by the following C. elegans cDNAs:  
YK4025.3, YK22a2.3, YK83f8.3, YK92f2.3, YK205b1.3,  
YK564g8.3, YK667a5.3, YK22a2.5, YK73d3.5, YK92f2.5,  
YK15a3.5, YK184c5.5, CESQ78f, YK205b1.5, YK399e12.5,  
YK400b1.5, YK402e5.5, YK555g2.3, YK555g2.5, YK619e7.3,  
YK667a5.5"

/codon\_start=1  
/product="Hypothetical protein Y51H7C.6a"  
/protein\_id="AAK39338.1"  
/db\_xref="GI:1377530"

## CDS

/translation="MPEPILHKSFLGSLPAGKNGVRHGEKPEKVEKQPFTRKIRE  
LRLLEIKREERIEKDIATILENTIDGSEONSFGLATYRLNNHMLVENSQKL  
TSALKINIVLADTISGRVSLADPAKTRVGCLOLAGDMRDGVCAGEJIDAIKSDFE  
TASOHIFRFLDQAVQIRFEKOKDADSTIRHSEVLSAKERSTIKLSRFSVO  
KGDVAEMORFKMRPLIHPREBGLQRYSVFNQKIDKLAEDNLAIKMGCGTDNRNV  
LYADLTFEFGVAEILISNLPVLEHSTGLKLDLDFETLQARIDEFRRLHEEDIR  
RRLSHFNRLVDVYIHQKKAADVEDADPMEIDALASEICMNTSAEMYWRFVSRIG  
KNEYIRSPSGDDEENEAROEERHRLKEKQKMDQLNRSRYGTMOELIGVCL  
LEHYMLKSVQKAIKSDYKEDAGLTSIVDVVYTIKRSIRASGVDSVCATIN  
NSATLIDIVVHGLRSTIOGGVTSNFASEAFYAOGKPKYKADAKOKEFLALN  
NSAKTSELLEILOGLITTEMAGVRPDKKLEKSTOIESAKKIASLAKHGYEEL  
FKSAFKNIKOGADPQOEDIRQMTQDVEYEAHDPEQETLQIDRLVEPELFA  
DNVQITLLTSSEIARQIESLACQFNRYALQIDRYRICALVTLNAGVSAEKV  
JIOGQIVSLNVTETIDEAMEVHNSKAMTSATITRLTLPVLRKVALARAPFPVAIK  
SIE"

join(2368..2797,2854..2979,3032..3261,4332..4496,  
4570..5103)  
/gene="Y51H7C.6"  
/note="coded for by the following C. elegans cDNAs:  
YK400b1.3, YK400b1.5"

/codon\_start=1  
/product="Hypothetical protein Y51H7C.6b"  
/protein\_id="AAK39339.1"  
/db\_xref="GI:1377531"

/translation="MPEPILHKSFLGSLPAGKNGVRHGEKPEKVEKQPFTRKIRE  
LRLLEIKREERIEKDIATILENTIDGSEONSFGLATYRLNNHMLVENSQKL  
TSALKINIVLADTISGRVSLADPAKTRVGCLOLAGDMRDGVCAGEJIDAIKSDFE  
TASOHIFRFLDQAVQIRFEKOKDADSTIRHSEVLSAKERSTIKLSRFSVO  
KGDVAEMORFKMRPLIHPREBGLQRYSVFNQKIDKLAEDNLAIKMGCGTDNRNV  
LYADLTFEFGVAEILISNLPVLEHSTGLKLDLDFETLQARIDEFRRLHEEDIR  
RRLSHFNRLVDVYIHQKKAADVEDADPMEIDALASEICMNTSAEMYWRFVSRIG  
KNEYIRSPSGDDEENEAROEERHRLKEKQKMDQLNRSRYGTMOELIGVCL  
LEHYMLKSVQKAIKSDYKEDAGLTSIVDVVYTIKRSIRASGVDSVCATIN  
NSATLIDIVVHGLRSTIOGGVTSNFASEAFYAOGKPKYKADAKOKEFLALN  
NSAKTSELLEILOGLITTEMAGVRPDKKLEKSTOIESAKKIASLAKHGYEEL  
FKSAFKNIKOGADPQOEDIRQMTQDVEYEAHDPEQETLQIDRLVEPELFA  
DNVQITLLTSSEIARQIESLACQFNRYALQIDRYRICALVTLNAGVSAEKV  
JIOGQIVSLNVTETIDEAMEVHNSKAMTSATITRLTLPVLRKVALARAPFPVAIK  
SIE"

## CDS

/gene="Y51H7C.9"  
/note="for a graphical representation of this gene see:  
http://www.wormbase.org/db/seq/sequence?name=Y51H7C.9;ccla  
ss=Sequence"  
complement(join(7941..7992,8063..8169,8301..8422,  
8469..8610,8657..8733,9106..9211,9269..9619,9674..9817,  
9865..9954,9999..10117,10177..10213))  
/gene="Y51H7C.9"

/note="coded for by the following C. elegans cDNAs:  
YK362g7.3, YK362g7.5"  
/codon\_start=1  
/product="Hypothetical protein Y51H7C.9"  
/protein\_id="AAK39340.1"  
/db\_xref="GI:1377532"

## Alignment Scores:

Pred. No.: 1.08e+03 Length: 68661  
Score: 84.50 Matches: 16  
Percent Similarity: 59.38% Conservative: 3  
Best Local Similarity: 50.00% Mismatches: 8  
Query Match: 52.16% Indels: 5  
DB: 3 Gaps: 1

US-09-997-610-2\_COPY\_18\_45 (1-28) x AC024805 (1-68661)

.. QY 2 G1YPROFALAHISPROAGPROGUGLUGVALG1YPROG1YALAPROG1YLEU 21

DB 66655 GGACCTCCAGGGGCTTCGACCTCCGGGGTACTGACCTGACCTCTGGGCTGACCA 66596

QY 22 PROGLINTYR-----Thrg1YGLUITE 28

DB 66595 CCCGAATACCGGTAGACACCGCTCACTGGCCAAATTA 66560

## RESULT 5

## AC006799/C

## LOCUS

## DEFINITION

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## COMMENT

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 2255: contig of 2255 bp in length  
\* 2256 2269: gap of unknown length  
\* 2270 5536: contig of 3267 bp in length  
\* 5537 5550: gap of unknown length  
\* 5551 23191: contig of 17641 bp in length  
\* 23192 23205: gap of unknown length  
\* 23206 60448: contig of 37243 bp in length  
\* 60449 60462: gap of unknown length  
\* 60463 112471: contig of 52009 bp in length  
\* 112472 112485: gap of unknown length  
\* 112486 188654: contig of 76169 bp in length  
\* 188655 188668: gap of unknown length  
\* 188669 278007: contig of 89339 bp in length.

## FEATURES

## source

## BASE COUNT

## ORIGIN

## Alignment Scores:

Pred. No.: 3.33e+03 Length: 278007  
Score: 84.50 Matches: 16  
Percent Similarity: 59.38% Conservative: 3  
Best Local Similarity: 50.00% Mismatches: 8  
Query Match: 52.16% Indels: 5  
DB: 2 Gaps: 1

US-09-997-610-2\_COPY\_18\_45 (1-28) x AC006799 (1-278007)

QY 2 G1YPROFALAHISPROAGPROGUGLUGVALG1YPROG1YALAPROG1YLEU 21

||||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 66577 GGACCTCCAGGGGTTCCTGGACTCACC66GGAAGTGGAACTCCTGGGATCACTGGACCA 66518

Qy 22 ProGlnTyT-----ThrGlyGuIle 28  
|||||  
|||::|||

Db 66517 CCCGAATACCGTAGAAGACCGTGCACTGCACATGGCCATTA 66482

RESULT 6  
AXI46430

LOCUS AXI46430 4428 bp DNA linear PAT 31-MAY-2001

DEFINITION Sequence 11 from Patent WO0134647.

ACCESSION AXI46430

VERSION AXI46430.1 GI:14284849

KEYWORDS .

SOURCE pig.  
ORGANISM Sus scrofa  
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Eumalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
1 (bases 1 to 4428)

REFERENCE 1 (bases 1 to 4428)  
AUTHORS Bell,M.P., Neff,T.B., Polarek,J.W. and Seeley,T.W.  
TITLE Animal collagens and gelatins  
JOURNAL Patent: WO 0134647-A 11 17-May-2001;  
FIBROGEN, INC. (US)

FEATURES  
source location/Qualifiers  
1..4428  
/organism="Sus scrofa"  
/db\_xref="taxon:9823" 910 t

BASE COUNT 994 a 1143 c 1381 g 910 t

ORIGIN

Alignment Scores:

Pred. No.: 133 Length: 4428

Score: 84.00 Matches: 16

Percent Similarity: 61.54% Conservative: 0

Best Local Similarity: 61.54% Mismatches: 10

Query Match: 51.85% Indels: 0

DB: 6 Gaps: 0

US-09-997-610-2-COPY\_18\_45 (1-28) x AXI46430 (1-4428)

Qy 1 AlaglyProProlahisProarArgProGluGluValGlyProProGlyAlaPProGly 20  
|||||  
|||||

Db 541 GCAGGTCCCCTGCCGCCACCTGTGTCCTGGTGTACTGTGTACTGTGTGTCCTGCTGTGT 600

Qy 21 LeuProGlnTyTrThrcily 26  
||| ||| |||  
||| ||| |||

Db 601 TCCTCAGATACCAAGG 618

RESULT 7  
AF017777

LOCUS AF017777 66669 bp DNA linear INV 04-AUG-1998

DEFINITION Drosophila melanogaster tweety (tty), flightless (fli), dodo (dod),  
penguin (pen), small optic lobes (sol), innocent bystander (iby),  
wacilin (wac), bobby sox (bbx), sluggish (slg), helicase (hlc),  
misto (mst), and la costa (lcs) genes, complete cds.

ACCESSION AF017777 U80043 U28044

VERSION AF017777.1 GI:3004652

KEYWORDS .

SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster.  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 66669)  
de Couet,H.G., Fong,K.S., Weeds,A.G., McLaughlin,P.J. and  
McKillop,G.L.  
TITLE Molecular and mutational analysis of a gelsolin-family member  
encoded by the flightless I gene of Drosophila melanogaster  
JOURNAL Genetics 141 (3), 1049-1059 (1995)  
MEDLINE 96129280  
PUBMED 8582612

REFERENCE 2 (bases 1 to 66669)  
AUTHORS Miklos,G.L., Yamamoto,M., Burns,R.G. and Malyszka,R.  
TITLE An essential cell division gene of Drosophila, absent from

JOURNAL	Saccharomyces, encodes an unusual protein with tubulin-like and myosin-like peptide motifs
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 94 (10), 5189-5194 (1997)
PUBMED	97289742
AUTHORS	3 (bases 1 to 66669)
TITLE	Maleszka, R., de Couet, H.G. and Miklos, G.L.
JOURNAL	Data transferability from model organisms to human beings: insights
MEDLINE	from the functional genomes of the flightless region of Drosophila
PUBMED	Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3731-3736 (1998)
REFERENCE	98188272
REFERENCE	9520435
REFERENCE	4 (bases 1 to 66669)
REFERENCE	Miklos, G.L.G., Kasprzak, A., Mason, J., de Couet, H.G., Hayward, D.,
REFERENCE	Hall, D. and Maleszka, R.
TITLE	Submitted (04-AUG-1997) Visual Sciences, Research School of
JOURNAL	Biological Sciences, The Australian National University, Sullivan
COMMENT	Creek Rd., Canberra, ACT 0200, Australia
COMMENT	On or before Apr 2, 1998 this sequence version replaced g1:209624,
COMMENT	g1:1052878.
FEATURES	Location/Qualifiers
source	1..66669
gene	/organism="Drosophila melanogaster"
gene	/strain="Cantons"
gene	/db_xref="taxon:7227"
gene	/chromosome="X"
gene	complement(1639..7694)
gene	/gene="fly"
gene	complement(join(1639..2707,2789..2800,3622..3791,
gene	3917..4680,4745..5230,5525..5809,7521..7694))
gene	/gene="fly"
gene	/product="fjolet1"
gene	complement(join(1767..2707,2789..2800,3622..3791,
gene	3917..4680,4745..5230,5525..5662))
gene	/gene="fly"
gene	/codon_start=1
gene	/product="lweely"
gene	/protein_id="AAC28400.1"
gene	/db_xref="GI:3004653"
gene	/translation="MGDYHEFMDQYKVPYATKLALAHYNTFFPKINSTFRPNDITY
gene	LESIGIGSVPAALLIYSLGLGLFTMTKRCDBRPPAHSTSLKVALSTIVTWCAG
gene	ILGIGYGGNDLHNEVLITAGKRVNLTITIRNDHLENTIRNRIPQVELADCA
gene	DQPSYNTAKSLKFLSVNITVOGVTATNAASDIRPLDGIISMTFTLRGOMDLIM
gene	PGYATVALLIYLCAYLVGVARRSHCALILFSCGLAVYSGMSGLYSSVAAG
gene	DLCSIPAEFLVSTARPDLPTNNVLIHTYCEPGHNPFGRIRESQNSINMRSAMATV
gene	IKSTIVLFSKSGICQIGAVNADINSESELDTQLNALDCKAHNNPAAARGICBEG
gene	ILGLVLIATISFTIAAILLTITWVDVSHITVITIRKRDAQVDEPSTISHAPQHOOM
gene	MNAATLTPLRRNNGHFSPPVTSGSHTLQHSKRSQHEHMAHAHIDQNMKMTHTLGR
gene	PSHNSPTHTMTPGNNAANAANAANAANMPPTQAAOQOQOQOQOQOQOQOQOQ
gene	OPITCNHPHOPHPHOPHSHSANAANAANOHAIAHQOQOQOQOQOQOQOQOQ
gene	HLIGQHPHOPHOPHOPHSHSANAANAANOHAIAHQOQOQOQOQOQOQOQOQ
gene	GRQIPAHOPHPSLHQOQOQOQPPPSOQOQOQOQOQOQOQOQOQOQOQHQQH
gene	HHQOQ
gene	KPSILNCPLEIPDAQPKVESIYEORQAHHQNVSK
gene	complement(5809..7521)
gene	/gene="fly"
gene	8586..13783
gene	/gene="f11"
gene	join(8368..8705,8869..12023,12260..12410,12533..13783)
gene	/gene="f11"
gene	/product="flightless"
gene	join(8652..8705,8869..12023,12260..12410,12533..12943)
gene	/gene="f11"
gene	/codon_start=1
gene	/product="flightless"
gene	/protein_id="AAC28407.1"
gene	/db_xref="GI:3004660"
gene	/translation="MSYLPVRGVDFPTKNDPSATPPSSROMSRVOMTIDRQIAT
gene	PEELGIQKLEHSLNNHNRLEKIFGLITGLSCRSLDLRHQKAKSGIPPELPIEL
gene	TTLDLSHNRKLEVEGELRAKNLTIVLNTSNNQIESITPPRLTHITDLEFLDSHNRLE
gene	TNPCEIRLISLSDHNPDELPTDRLPDLPSGLVLYKSGTORLTLLNPPSIDSL
gene	ANLPEQDLSHNSLPKLPDCYVNVVTLVRNLNSDNETFELTGVGVLMQRTLSNRQ

IVALTEAAALCKLPKLRLLLVNDKRLNFEPIDPSIGKGLAEVFSAAANNLLEAVPELCR  
 GCAALNOLGKRLITLDPADHILEGLGGLDOLRRNPPELVMPKPEASARLSLEIYR  
 DESGLOALRIGAGAAVPSMPSPSAPTPKSDTAKIRLRRRPRESEGDODAAKVALKMGDVA  
 KDKDEAKVAPEDGKPSLPKRRWDESKLEKPOLDSKFEFKDGLDGPULTEIENFL  
 PNKLEEVHGKVFEGDQCYIVLTKTKFDDGLDLMELFEMIAGNATLDRKACAAIAHVL  
 NFNGLFARCRVAREEGDSEQFLSLFEFTEVHIYIEGGFRATGYTTEEMIHITRLIVH  
 AVGAIHIEPVAPIATSLDPRHAFVLDTGTHIYIMGSRKNTLNSKARIAMAEISKI  
 BRKNCKCELEFVAREEGESAEPMQGLGCTSEEDADAPRPKEHNPEDVOPAPORLYOVLG  
 MGYLEIPOVELPEQKLCNTLNSKHVYILDCYTDFVFGKSKITPLVAAAVKLSREL  
 FNMDRPDIALVMVPEGNEMOIFRTKTAGDVEVAVOTFRAKSVARTAGVALTQWAR  
 QOETRTDIALALEMPROSAMPLAEALQLEENMYDLEMMEAFLVLEKKKFEVRLPEBELGR  
 FYETGCVLALCYPCIPRIEPENGSESDGNAPDAYSKSSANNOPDEDEIOCVVFMQGR  
 AGNMGLMFEFTGLKPKFKAMGEELEUVRIPOOQNLNMFHFKRKLFIHNGRRDKA  
 HTAKSKSVPEFPHLRNSGALITLTRLIOINPAVHINSAPCYLIHNPETEDDSOGIV  
 YVWISKACNEBAKLVODIAEQMFSNPVSLQIILEGDEPENEFVVALGKRPITDIA  
 EVMATRLRSCSNERGYTVAEKCADFCODDLADIDIMILONGEHVFLMGPCESEVE  
 VKLASQVYIQHMRIRKQPERPRKLFTLMKKNESRRFTKCFHGMASERKVVYL"  
 14131. .15603  
 /gene="dod"  
 /name="dod"  
 join(14131. .14252,14364. .14696,15066. .15603)  
 /gene="dod"  
 /product="dodo"  
 join(14195. .14252,14364. .14696,15066. .15175)  
 /gene="dod"  
 /codon\_start=1  
 /product="dodo"  
 /protein\_id="AAC28408.1"  
 /db\_xref="GI:3004651"  
 /translaltion="MPDAEQLPDGKEKRTSRSTGMSYVLMNYTERESQNDPTEPAKKT  
 GGSAGGDAPDEHCLHLVYKHSRSPSSREARNTIRTEADALILEVYRNKIVQO  
 BATPELIRASVSDCSSARKGDLGKFGNGQMAPEADAFKILNVNQSLGYSDSGLH  
 ILIRKA"  
 complement(15247. .18587)  
 /gene="pen"  
 /name="pen"  
 complement(join(15247. .15493,15707. .16182,16281. .16683,  
 16752. .17454,18055. .18587))  
 /gene="pen"  
 /product="penguin"  
 complement(join(15308. .15493,15707. .16182,16281. .16683,  
 16752. .17454,18055. .18500))  
 /gene="pen"  
 /note="contains pumilio repeats"  
 /codon\_start=1  
 /product="penguin"  
 /protein\_id="AAC28403.1"  
 /db\_xref="GI:3004656"  
 /translaltion="MYSSGEPKGPANVANTASPLKHAHSSGKRANGOKFKPGAGARK  
 FDKPGGAGKSVFKPGAGGAKKFKDSGFGFKPKPSGATGKRLGSKMPKPGKQPOT  
 OAPAGEKQDMNKKFKKKDKLTKRKSADRYEISKEANOJHEKIRGRTENKQKLYE  
 QIYKALNGDTISKVVKAHDPTRVYQSMLKATASPLRAEISEKLLPTEVEMQSYAO  
 FCVQMLKLYGAATKAKLVDSLGHIVRLAGHSISGLSDSNYOSATNORJMYOEY  
 YGLDRKRRKSDLEETVTAFAALVPHMLSTKEGSEAVAYCFKSPKRNRAIINKHE  
 YLRACDDEEKLKEETVTAFAALVPHMLSTKEGSEAVAYCFKSPKRNRAIINKHE  
 LKLIHNHGHVFLISLINALDDTPKAKKAIYDHLDGLKALSDSPGRCVRLYQVAP  
 GDTGTFHEPRTVPEGLAFGKKKEKLEKELBDETPRPAOSTAEDAPWLSNHTG  
 LVTDGIHLHIOESTERKASALAOVVOPPEWTSIDAAGPOPODKKRVHNVESALIAQ  
 ATKQRKRLAAYPSSDDEDEDDEDEDEDEDEKDEKQEAADAEPRVYKAKKSPK  
 PKAKEEPLAEPVLSGIEEAGNHIVYKLIKNDGKEGTFPSQOLLQONLSVPLKMLG  
 VNRAEFLVLTKEECPPALLDCKKAIAMERSISQILADRKTPGAKILAAKIDICK"  
 19098. .26653  
 /gene="sol"  
 /name="sol"  
 join(19098. .19602,20609. .21234,21502. .22414,22843. .23171,  
 23238. .24826,24899. .25405,25465. .25731,25803. .25945,  
 26004. .26653)  
 /gene="sol"  
 /product="small optic lobes"  
 join(19361. .19602,20609. .21234,21502. .22414,22843. .23171,  
 23238. .24826,24899. .25405,25465. .25731,25803. .25945,  
 26004. .26181)  
 /gene="sol"  
 /codon\_start=1  
 /product="small optic lobes"  
 /protein\_id="AAC28409.1"

```

/adb.xref="Gi.3004662"
/translation="MCTISSVLOMCSCKCMTIPTEJIKCFNGCTGVKVCYPR00000H
RSSSTIRASTADADLEBOAEKOEGBDEKKEGRAVAVASEKHYYKILRECLKRPON
SONLPAWCVCDEETRYIKSSIELYRHFSPALNRRWVCAAGSTDNSSVTHCLICDT
VSLYAPLRYKDAIADROGDLAGISGNLEPLADHPRHHNNHYLHOELDEOHONHDL
SOULKHNLKGRSAGSGSGSGSGSLGRPLSTLADISASGSKHICVANNOSKOTI
FNLPIKRAPOLGIPVAVACSNSRFLIADTPFCRRKONNNKNNKHYVREAGKRRK
YNTTITLSAKAKDACHQMKRPLYVNLNLNLOOEPOKSPANKPOLKOTQREBA
AVSMNPOTPTIPRNGEYIAVNEWSEPAASSVSSSSSSHHNNHNSNSNSNSNTI
NNSSSSSGSNKLYENECVLAQOOLAAQOOLAAQOAAATVVALIASPSAAMAPET
ATMPTIYAQVOKHKLKKKQOJASESTNNNGEELIDAVSESLTGLGTSTGGSGA
SESEJOYEHEHSTYAKWKQGPRAKTEPIMDDPSSSGLSCAAAGAATGAGIAAAY
GAAASRHNKKTOLGONSRSKMWICIKCSYAYRMLYQTEMKAEAEQOQOOLQLO
QOQOQOHHNNHHNNHHNNHLOQOQAEAPRDEPPTCKCTLYNLSTACAVCGGSKLXIS
IEMTLRKBEFTCSHCOTLNSLHSPYCSAKHROPOLSMAEVAERDGOSEYED
DAAVGGGSGAHOSSGANEVKAAPTALNPLTVSALPMYLPLOTSTPAGIGRSRSP
RMLLPLOOORNSSSGAIPKRRHSTGSLVPNIRISIAGLANTNLQOOGVGSASVS
AASGGGAGAVGASTSSKWCQCPACTYDNCAAVNDICSGPSGLASVYGLAGRS
VYRALTPADIRKESKLMENLRQLEBTEALTKWONIYOICDNSLELYDSSFPAKPSL
VYPAASGAGBNVYVOWRPHREINDCOGAAVPAWVFPTPLPSDICQULANCMYLSL
AVLAEREDLVKLEVLTKETICOGGAYQRLCKDKMTVVALPESDPCRKCHLYVQAK
RQOLWDLKAVAKIHGCEYALVSGAIPKRTTGLTIPADSLIPLOASSPMESEDL
DKRLIAPOLLSRCVRLPAGASGGGKMGADDEEYOGKGPRAHAYSLVLYKIOGHR
LLALRNPNGHYNRGDMSDSSLTMTDRLPALRPHAKSEBVNLSPEYDLYTDCIDI
CKVRSQWNEVRLQTLQPLCISICVLTLVEPTEAETTLPEQESORNSKSORSOULD
VYLFERTSPAAREIGRLVESHKSROVRFVCHMLMEIDLALVLCANHHNTLIEBPA
YOQOCLIAHSKRLEAVOISPSPLHLADISITTLTGORHGBREGMTAYVLYTKGA
GLVYVNEHRENNHNYKCDOSGSYVNVSRGELQYDVSYPLODROYIYLTOLESGG
GESIAHRUTLRNLANSKRLHDMWPPGATNCPPIENVAGLHAPRLIT"
27035. .30414

```

Alignment Scores:			
Pred. No.:	1.17e+03	Length:	66669
Score:	84.00	Matches:	14
Percent Similarity:	66.67%	Conservative:	0
Best Local Similarity:	66.67%	Mismatches:	7
Query Match:	51.85%	Indels:	0
DB:	3	Gaps:	0
US-09-197-610-2_COPY_18_45 (1-28) x AF017777 (1-66669)			
QY	2	GlyprophalaHsiproHsprogrogluGlvalGlyproGlyAlaProGlyUeu 21	
Db	65789	GGGCTCTCTGGGCGACCTGGCTCTCTGGGCTCTGGGCTCTGGCTCTGGCTCTGGGCT 65848	
QY	22	Pro 22	
Db	65849	CCT 65851	
RESULT 8			
AC013980			
LOCUS			
DEFINITION			
AC013980 87089 bp DNA linear HTG 16-NOV-1999			
Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, In ordered			
pieces.			
AC013980			
AC013980.1 GI:6437355			
HTG: HTGS_PHASE2.			
Drosophila melanogaster.			
Drosophila melanogaster			
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
Ephydroidea; Drosophilidae; Drosophila.			
1 (bases 1 to 87089)			
Adams,M. and Venter,J.C.			
Direct Submission			
Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,			
Rockville, MD, USA			
This sequence was identified as CDN:10212172 by the submitter.			
For further information on this sequence e-mail to fly@celera.com.			
* NOTE: This is a 'working draft' sequence.			
* This sequence will be replaced			
* by the finished sequence as soon as it is available and			
* the accession number will be preserved.			

FEATURES	Location/Qualifiers
source	1..87089
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
BASE COUNT	23787 a 19214 c 18797 g 25291 t
ORIGIN	
Alignment Scores:	
Pred. No.:	1.45e+03
Score:	84.00
Percent Similarity:	66.67%
Best Local Similarity:	66.67%
Query Match:	51.858
DB:	2
	Gaps: 0
US-09-997-610-2_COPY_18_45 (1-28) x AC013980 (1-87089)	
QY	2
	ProbaHahsProrAhrProProgiUgUaValGlyProrOgYAlaProgYleu 21
	ProbaHahsProrAhrProProgiUgUaValGlyProrOgYAlaProgYleu 21
Db	51838
	GGGCTCTCTGGGCAATCCCTGCTCCTGGGCGCTCCTGGTCTCCTGGTCTCCTGGGCGCT 51897
QY	22
	Pro 22
Db	51898
	CCT 51900
RESULT 9	
AC092494	171548 bp DNA linear INV 12-JUL-2001
LOCUS	Drosophila melanogaster, chromosome X, region 20B-20C, BAC clone
DEFINITION	BACR23118, complete sequence.
AC092494	
VERSION	AC092494.1 GI:14701995
KEYWORDS	HTG.
SOURCE	Drosophila melanogaster.
ORGANISM	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
	Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 171548)
AUTHORS	Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Man,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Bazon,J., Beeson,K.Y., Busam,D.A., Carlson,J.W., Center,A., Chape,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorselt,V., Doup,L.E., Doyle,C., Dresnek,D., Fairlan,D., Fierlier,S., Frise,E., Galle,R.F., Gary,N.S., George,R.A., Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J., Idegami,C., Jallai,M., Kruse,D., Li,P., Mettel,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Pacib,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Phouenavong,S., Piltman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Sylvestre,R., Tector,C., Williams,S.M., Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
TITLE	Sequencing of Drosophila chromosome X, region 20B-20C
JOURNAL	Unpublished
AUTHORS	2 (bases 1 to 171548)
	Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Man,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Bazon,J., Beeson,K.Y., Busam,D.A., Carlson,J.W., Center,A., Chape,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorselt,V., Doup,L.E., Doyle,C., Dresnek,D., Fairlan,D., Fierlier,S., Frise,E., Galle,R.F., Gary,N.S., George,R.A., Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J., Idegami,C., Jallai,M., Kruse,D., Li,P., Mettel,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Pacib,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Phouenavong,S., Piltman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Sylvestre,R., Tector,C., Williams,S.M., Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
TITLE	Direct Submission
JOURNAL	Submitted (12-Jul-2001) Berkeley Drosophila Genome Project, MS 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, US
COMMENT	Sequence submitted by:

	Berkeley Drosophila Genome Project Lawrence Berkeley National Laboratory, MS 64-121 Berkeley, CA 94720
	This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site ( <a href="http://www.fruitfly.org/sequence/">http://www.fruitfly.org/sequence/</a> ) or send email to <a href="mailto:bdp@fruitfly.berkeley.edu">bdp@fruitfly.berkeley.edu</a> .
FEATURES	Location/Qualifiers
SOURCE	1..171548 /organism="Drosophila melanogaster" /strain="y; cn by sp" /db_xref="taxon:7227" /chromosome="X" /map="20B-20C" /clone="BACR23118 (D1450)" /clone.lib="RPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBAC5.6)"
BASE COUNT	47824 a 36178 c 37366 g 50180 t
ORIGIN	
Alignment Scores:	
Pred. No.:	2.49e+03 Length: 171548
Score:	84.00 Matches: 14
Percent Similarity:	66.67% Conservative: 0
Best local Similarity:	66.67% Mismatches: 7
Query Match:	51.85% Indels: 0
DB:	3 Gaps: 0
US-09-997-610-2_COPY_18_45 (1-28) x AC092494 (1-171548)	
OY	2 GLYPProFALahisPcArGPrOgLoLuVaLGLYPProGlyAapProGlyLeu 21 
Db	51719 GGTCCTCGTGGCATCTGCTCTCTGGGCCCTGTCTCTCTCTCTCTGGCCCT 51778
OY	22 Pro 22 
Db	51779 CCF 51781
RESULT 10	
LOCUS	AC011251
DEFINITION	AC011251 173508 bp DNA linear INV 07-JUN-2001
ACCESSION	Drosophila melanogaster, chromosome X, region 19F-20A, BAC clone
VERSION	BACR09F10, complete sequence.
KEYWORDS	AC011251
SOURCE	AC011251.5 GI:14327739
ORGANISM	HMG. Drosophila melanogaster. Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Phytotroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 173508) Celniker,S.E., Adams,M.D., Krommler,B., Tyler,D., Wan,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A., Carlson,J.W., Center,A., Chape,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dotselt,V., Doup,L.E., Doyle,C., Dresner,D., Fattan,D., Ferriter,S., Frisic,E., Galle,R.F., Garg,N.S., George,R.A., Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J., Idegam,C., Jallali,M., Kruse,D., Li,P., Mattei,B., Moshefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Packeb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Phoumenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Svrlidskas,R., Tector,C., Williams,S.M., Zaveril,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
TITLE	JOURNAL
REFERENCE	Sequencing of Drosophila chromosome X, region 19F-20A
AUTHORS	Unpublished 2 (bases 1 to 173508) Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazek,J.R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,



\* 36719 36818: gap of unknown length  
\* 36819 36835: contig of 1537 bp in length  
\* 38356 38455: gap of unknown length  
\* 38456 39544: contig of 1089 bp in length  
\* 39545 39644: gap of unknown length  
\* 39645 42096: contig of 2452 bp in length  
\* 42097 42196: gap of unknown length  
\* 42197 44438: contig of 2242 bp in length  
\* 44439 44538: gap of unknown length  
\* 44539 46550: contig of 2112 bp in length  
\* 46551 46751: gap of unknown length  
\* 46751 48936: contig of 2186 bp in length  
\* 48937 49035: gap of unknown length  
\* 49037 52140: contig of 3104 bp in length  
\* 52141 52240: gap of unknown length  
\* 52241 54747: contig of 2507 bp in length  
\* 54748 54848: gap of unknown length  
\* 54848 59017: contig of 4170 bp in length  
\* 59018 59118: gap of unknown length  
\* 59118 62801: contig of 3684 bp in length  
\* 62802 62901: gap of unknown length  
\* 62902 66517: contig of 3616 bp in length  
\* 66518 66618: gap of unknown length  
\* 66618 70275: contig of 3657 bp in length  
\* 70275 70375: gap of unknown length  
\* 70375 75340: contig of 5165 bp in length  
\* 75340 75640: gap of unknown length  
\* 75640 79310: contig of 3670 bp in length  
\* 79310 84690: gap of unknown length  
\* 84690 84790: gap of unknown length  
\* 84790 88910: contig of 5021 bp in length  
\* 88910 89911: gap of unknown length  
\* 89911 95503: contig of 5593 bp in length  
\* 95504 95604: gap of unknown length  
\* 95604 100929: contig of 5325 bp in length  
\* 100929 101029: gap of unknown length  
\* 101029 105513: contig of 4485 bp in length  
\* 105514 105613: gap of unknown length  
\* 105614 110389: contig of 4776 bp in length  
\* 110390 110489: gap of unknown length  
\* 110490 114852: contig of 4363 bp in length  
\* 114853 114953: gap of unknown length  
\* 114953 120726: contig of 5774 bp in length  
\* 120727 120826: gap of unknown length  
\* 120827 127059: contig of 6233 bp in length  
\* 127060 127159: gap of unknown length  
\* 127160 133607: contig of 6448 bp in length  
\* 133608 133707: gap of unknown length  
\* 133708 138628: contig of 5921 bp in length  
\* 138629 139728: gap of unknown length  
\* 139729 147785: contig of 8057 bp in length  
\* 147786 147885: gap of unknown length  
\* 147886 154861: contig of 6976 bp in length  
\* 154862 154961: gap of unknown length  
\* 154962 160982: contig of 6021 bp in length  
\* 160983 161082: gap of unknown length  
\* 161083 167777: contig of 6694 bp in length  
\* 167778 167876: gap of unknown length  
\* 167877 175123: contig of 7247 bp in length  
\* 175124 175223: gap of unknown length  
\* 175224 183464: contig of 8241 bp in length  
\* 183465 183564: gap of unknown length  
\* 183565 191381: contig of 7816 bp in length  
\* 191382 191480: gap of unknown length  
\* 191481 200932: contig of 9452 bp in length  
\* 200933 201032: gap of unknown length  
\* 201033 207587: contig of 6535 bp in length  
\* 207588 207687: gap of unknown length  
\* 207688 216967: contig of 9280 bp in length  
\* 216968 217067: gap of unknown length  
\* 217068 229646: contig of 12579 bp in length  
\* 229647 229746: gap of unknown length

\* 229747 241492: contig of 11746 bp in length  
\* 241493 241592: gap of unknown length  
\* 241593 256417: contig of 14825 bp in length.  
FEATURES  
source  
1..256417  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="16"  
/clone="RP23-109A22"  
/clone\_11b="RPC1 mouse BAC library 23"  
BASE COUNT 67041 a 56625 c 61089 g 63261 t 8401 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 3.44e+03 Length: 256417  
Score: 84.00 Matches: 15  
Percent Similarity: 68.18% Conservative: 0  
Best Local Similarity: 68.18% Mismatches: 7  
Query Match: 51.65% Indels: 0  
DB: 2 Gaps: 0  
US-09-997-610-2\_copy\_18\_45 (1-28) x AC079422 (1-256417)  
QY 4 PROAHHSPTGARGPRGPGSLGVLVGLPRGPGLYALPRGPGLYLEPRGGIN 23  
Db 100552 CCGCGCCCTGCCCCCGGACCGCGCGCCCGCGCCCGCGCCCTGCCCA 100611  
QY 24 TYRTHR 25  
Db 100612 ACNACG 100617  
RESULT 12  
AE003568 292911 bp DNA linear INV 05-OCT-2000  
LOCUS Drosophila melanogaster genomic scaffold 14200013386037 section 5  
DEFINITION of 5, complete sequence.  
ACCESSION AE003568 AE002620  
VERSION AE003568.2 GI:10726967  
KEYWORDS HTG.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 292911)  
Adams,M.D., Celiker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,  
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,  
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,  
Sutton,G.G., Wortman,J.R., Vandeil,M.D., Zhang,Q., Chen,L.X.,  
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,  
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor  
Miklos,G.L., Abrell,J.F., Agbayani,A., An,H.J.,  
Andrews-Pfankoch,C., Baldwin,D., Bailew,R.M., Basu,A.,  
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,  
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borokova,D.,  
Botchan,M.R., Bouck,J., Brokstein,P., Brothier,P., Butts,K.C.,  
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,T.,  
Cherry,J.M., Cavley,S., Dahike,C., Davenport,L.B., Davies,P., de  
Pablos,B., Delcher,A.D., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,  
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,  
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,  
Fleischmann,W., Foster,C., Gabriellian,A.E., Garg,N.S.,  
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gottlieb,J.H., Gu,Z.,  
Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,  
Hernandez,J.R., Houck,J., Hostlin,D., Houston,K.A., Howland,T.J.,  
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,  
Kennison,D.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,  
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,  
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,  
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,  
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,  
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,  
Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Piltman,G.S., Pan,S.,

[illegible][illegible]

CDS

/db\_xref="FLYBASE:FBgn0011718"  
/db\_xref="FLYBASE:FBgn001170"  
complement(join(65188..65380,65443..66216,66361..66537,  
66595..66745,66950..67802,67867..68234,68300..70124,  
70196..70784,72001..72212))  
/gene="CG1718"  
/note="CG1718 gene product"  
/codon\_start=1  
/protein\_id="AA550837.1"  
/db\_xref="GI:7295524"  
/db\_xref="FLYBASE:FBgn0011718"  
/db\_xref="FLYBASE:FBgn001170"  
/translation="MAKVTNMDEKLVLLKNNFTLQNHKMGVLEIVLPATFSLILV  
VRIIVDEQKGVKRYNEQNTLDNLQKNGSKFEPLICTSPVNPVLKKLVEAMOS

## Alignment Scores:

Pred. No.: 3,83e+03 Length: 292911  
Score: 84.00 Matches: 14  
Percent Similarity: 66.67% Conservative: 0  
Best Local Similarity: 66.67% Mismatches: 7  
Query Match: 51.85% Indels: 0  
DB: 3 Gaps: 0

US-09-997-610-2\_COPY\_18\_45 (1-28) x AE003568 (1-292911)

QY 2 GLYPROPTOLAHISPTARGPROBGLUGLVAIGLYPROBGLYALAPROGLYLEU 21

Db 168892 GGTCTCTCGGCGATCGTCTCTCGGCGCTCTGCTCTGCTCTGCTCTGCGCT 168951

QY 22 Pro 22

Db 168952 CCT 168954

## RESULT 13

LOCUS AR014097 186 bp DNA linear PAT 05-DEC-1998  
DEFINITION Sequence 30 from patent US 5773249.  
ACCESSION AR014097  
VERSION AR014097.1 GI:3971551

KEYWORDS .

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 186)  
AUTHORS Cappello,J. and Ferrari,F.A.  
TITLE High molecular weight collagen-like protein polymers  
JOURNAL Patent: US 5773249-A 30 30-JUN-1998;  
FEATURES Location/Qualifiers

source 1..186

BASE COUNT 21 a 71 c 67 g 27 t

ORIGIN

## Alignment Scores:

Pred. No.: 12.7 Length: 186  
Score: 83.00 Matches: 15  
Percent Similarity: 61.54% Conservative: 1  
Best Local Similarity: 57.69% Mismatches: 10  
Query Match: 51.23% Indels: 0  
DB: 6 Gaps: 0

US-09-997-610-2\_COPY\_18\_45 (1-28) x AR014097 (1-186)

QY 2 GLYPROPTOLAHISPTARGPROBGLUGLVAIGLYPROBGLYALAPROGLYLEU 21

Db 37 GGTCTCTCGGCGGCGTCTCTCAAGACCGCGCTGCTCTGCTCTGCGGCGACTG 96

QY 22 ProGLNTYrThrGLYGLU 27

Db 97 CCAGGCCCGAAGGCGAT 114

## RESULT 14

AR117066

LOCUS AR117066 186 bp DNA linear PAT 16-MAY-2001  
DEFINITION Sequence 62 from patent US 6140072.  
ACCESSION AR117066  
VERSION AR117066.1 GI:14097972

KEYWORDS .

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 186)  
AUTHORS Ferrari,F.A. and Cappello,J.  
TITLE Functional recombinantly prepared synthetic protein polymer  
JOURNAL Patent: US 6140072-A 62 31-OCT-2000;  
FEATURES Location/Qualifiers

source 1..186

BASE COUNT 21 a 71 c 67 g 27 t

ORIGIN

## Alignment Scores:

Pred. No.: 12.7 Length: 186  
Score: 83.00 Matches: 15  
Percent Similarity: 61.54% Conservative: 1  
Best Local Similarity: 57.69% Mismatches: 10  
Query Match: 51.23% Indels: 0  
DB: 6 Gaps: 0

US-09-997-610-2\_COPY\_18\_45 (1-28) x AR117066 (1-186)

QY 2 GLYPROPTOLAHISPTARGPROBGLUGLVAIGLYPROBGLYALAPROGLYLEU 21

Db 37 GGTCTCTCGGCGGCGTCTCTCAAGACCGCGCTGCTCTGCTCTGCGGCGACTG 96

QY 22 ProGLNTYrThrGLYGLU 27

Db 97 CCAGGCCCGAAGGCGAT 114

## RESULT 15

LOCUS I06482 186 bp DNA linear PAT 02-DEC-1994  
DEFINITION Sequence 19 from Patent WO 9005177.  
ACCESSION I06482  
VERSION I06482.1 GI:589523

KEYWORDS .

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 186)  
AUTHORS Cappello,J., Ferrari,F.A. and CO7H15.  
TITLE FUNCTIONAL RECOMBINANTLY PREPARED SYNTHETIC PROTEIN POLYMER  
JOURNAL Patent: WO 9005177-A 19 17-MAY-1990;  
FEATURES Location/Qualifiers

source 1..186

BASE COUNT 21 a 71 c 67 g 27 t

ORIGIN

## Alignment Scores:

Pred. No.: 12.7 Length: 186  
Score: 83.00 Matches: 15  
Percent Similarity: 61.54% Conservative: 1  
Best Local Similarity: 57.69% Mismatches: 10  
Query Match: 51.23% Indels: 0  
DB: 6 Gaps: 0

US-09-997-610-2\_COPY\_18\_45 (1-28) x I06482 (1-186)

QY 2 GLYPROPTOLAHISPTARGPROBGLUGLVAIGLYPROBGLYALAPROGLYLEU 21

Db 37 GGTCTCTCGGCGGCGTCTCTCAAGACCGCGCTGCTCTGCTCTGCGGCGACTG 96

QY 22 ProGLNTYrThrGLYGLU 27

Db 97 CCAGGCCCGAAGGCGAT 114

Thu Feb 20 14:55:13 2003

us-09-997-610-2\_copy\_18\_45.rge

Page 13

Search completed: February 19, 2003, 23:10:55  
Job time : 581.136 secs

---

**THIS PAGE BLANK (USPTO)**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

2084.671 Million cell updates/sec

Description
BM967732 LM24HW01
BE757276 211669 M
BE757275 211668 M

AL320996 Tetraodo

**BLOSUM62**

Xgapop	10.0	,	Xgapext	0.5
Ygapop	10.0	,	Ygapext	0.5
Fgapop	6.0	,	Fgapext	7.0
Delop	6.0	,	Delext	7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Command line parameters:

```
MODEL=Hrime+P2n.model -DEV=x1h
-Q/cgcn2_1/USPO.spool/US050997610/runcat.10022003.160823.23843/app.qvec.fasta_1.1633
-DB=EST -GEMT=fastap -SUFFIX=first -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=humano4.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFM=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US050997610.ecgn.1.1871.gnumat.10022003.160823.23843 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAPP -LARGESQUEERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database : EST: \*

1:	em_estbn:*
2:	em_estbn:*
3:	em_estbn:*
4:	em_estbn:*
5:	em_estbn:*
6:	em_estbn:*
7:	em_estbn:*
8:	em_estbn:*
9:	em_estbn:*
10:	em_estbn:*
11:	em_estbn:*
12:	em_estbn:*
13:	em_estbn:*
14:	em_estbn:*
15:	em_estbn:*
16:	em_estbn:*
17:	em_estbn:*
18:	em_estbn:*
19:	em_estbn:*
20:	em_estbn:*
21:	em_estbn:*
22:	em_estbn:*
23:	em_estbn:*
24:	em_estbn:*
25:	em_estbn:*
26:	em_estbn:*
27:	em_estbn:*

## ALIGNMENTS

No.	Score	Query	Length	DB	ID	Description
1	306.5	43.0	499	14	BM67732	BM67732 LM24HM013
2	228.5	32.1	526	12	BE757276	BE757276 211669 MA
3	226.5	31.8	504	12	BE757275	BE757275 211668 MA
4	181	25.4	663	13	BJ068972	BJ068972 BJ068972
5	165.5	23.2	1092	17	CNS05408	AL240996 Tetradion
6	158.5	22.3	512	13	BJ494317	BJ494317 BJ494317
7	158	22.2	587	10	AV93019	AV93019 AV93019
8	157.5	22.1	689	10	BB318780	BB318780 BB318780
9	157.5	22.1	2462	11	AK018742	AK018742 Mus muscu
10	157	22.1	662	14	BQ419957	BQ419957 faa12c05.
11	155.5	21.8	695	9	AV243093	AV243093 AV243093
12	154.5	21.7	921	9	AA673154	AA673154 vef0g07.r
13	152.5	21.4	468	10	AP988554	AP988554 uq08h07.y
14	152.5	21.4	552	10	BE625509	BE625509 uu19e06.y
15	152.5	21.4	633	10	BB625284	BB625284 BB625284
16	152.5	21.4	707	12	BG872725	BG872725 602793813
17	152.5	21.4	921	11	AK003138	AK003138 Mus muscu
18	151.5	21.3	460	9	AI332213	AI332213 fae7d02.y
19	150.5	21.1	508	9	AA493090	AA493090 v177g11.r
20	150.5	21.1	589	14	BQ193813	BQ193813 UT-R-CM1-
21	150.5	21.1	885	9	AI132311	AI132311 mJ87e03.y
22	149	20.9	542	13	BI760895	BI760895 603043764
23	149	20.9	989	14	BQ900894	BQ900894 AGENCOURT
24	147.5	20.7	856	14	BQ876848	BQ876848 AGENCOURT
25	146.5	20.6	607	12	BG691150	BG691150 AAG0196 BA
26	146	20.5	585	10	AM913280	AM913280 ufe1sb09.y
27	145.5	20.4	580	12	BG691162	BG691162 340211 BA
28	145.5	20.4	582	10	BE588478	BE588478 196950 BA
29	145.5	20.4	1029	17	CNS024P5	AL178466 Tetradion
30	144.5	20.3	453	10	BE126709	BE126709 DEPK0457
31	144.5	20.3	639	13	BJ090235	BJ090235 BQ090235
32	143	20.1	430	10	AA453561	AA453561 zeh10104
33	142.5	20.0	693	9	AA060543	AA060543 mj64c11.r
34	141.5	19.9	890	14	BO737531	BO737531 AGENCOURT
35	141	19.8	683	10	AM029321	AM029321 wx09a01.x
36	140.5	19.7	529	12	BF074104	BF074104 221439 MA
37	140	19.7	482	14	BQ260214	BQ260214 faa09b03
38	140	19.7	542	14	BO449231	BO449231 faa44f11.
39	140	19.7	563	14	BQ285664	BQ285664 faa50f03.
40	140	19.7	566	14	BQ258785	BQ258785 faa24f09.
41	140	19.7	580	13	BI477496	BI477496 da142g12.
42	140	19.7	587	14	BQ420027	BQ420027 faa13b04.
43	140	19.7	597	14	BO285786	BO285786 faa29n11.
44	140	19.7	598	14	BO420183	BO420183 faa15a02.
45	140	19.7	626	14	BQ260132	BQ260132 faa08b03.

RESULT 1	
BM967732	
LOCUS	499 bp mRNA linear EST 20-MAR-2002
DEFINITION	LM-24HW0134 Bos taurus LM-24-HW CDNA library Bos taurus CDNA clone
ACCESSION	BM967732
VERSION	BM967732.1 GI:19561919
KEYWORDS	EST.
SOURCE	cow.
ORGANISM	Bos taurus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
AUTHORS	1 (bases 1 to 499) Yoon,D.H., Jang,Y.S., Kim,T.H., Park,E.W., Lee,H.K., Chung,E.R.,



```
RESULT 3
BE757275          504 bp  mRNA  linear  EST 25-APR-2001
LOCUS             BE757275
DEFINITION        211668 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION         BE757275
VERSION           BE757275.1  GI:10171267
KEYWORDS          EST.
SOURCE            cow.
  ORGANISM        Bos taurus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
    Bovidae; Bovinae; Bos.
  1 (bases 1 to 504)
REFERENCE          Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
AUTHORS            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,
                  'G.L., Heaton,M.P., Laegreid,W.W., Kohrer,G.A., Chitko-McKown,C.G.,
                  Perlea,G., Holt,I., Karameycheva,S., Liang,F., Quackenbush,J. and
                  Keefe,J.W.
                  Sequence evaluation of four pooled-tissue normalized bovine cDNA
                  libraries and construction of a gene index for cattle
TITLE              Genome Res. 11 (4), 626-630 (2001)
JOURNAL            21180013
MEDLINE            Contact: Smith TPL
COMMENT            USDA, ARS, US Meat Animal Research Center
                  PO Box 166, Clay Center, NE 68933-0166, USA
                  Tel: 402 762 4366
                  Fax: 402 762 4390
                  Email: smithhemall@marc.usda.gov
                  Single pass sequencing. Bases called and alt-trimmed with phred
                  v0.980904.e. Vector identified by cross_match with the -linscore 18
                  and -minmatch 12 options.
                  PCR primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTTCCAGTCACGACG
Plate: 63 row: G column: 13
Seq primer: ATTTAGTCACATATAG.
FEATURES
  source            Location/Qualifiers
                  1..504
                  /organism="Bos taurus"
                  /db_xref="taxon:9913"
                  /clone_lib="MARC 2BOV"
                  /tissue_type="Pooled"
                  /lab_host="DH10B"
                  /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                  library made from pooled tissue from testis, thymus,
                  semitendinosus muscle, longissimus muscle, pancreas,
                  adrenal, and endometrium."
BASE COUNT         99 a 131 c 146 g 128 t
ORIGIN
Alignment Scores:
Pred. No.:         1,25e-14      Length:      504
Score:             226.50        Matches:     51
Percent Similarity: 55.86%      Conservative: 11
Best Local Similarity: 45.95%   Mismatches:  38
Query Match:       31.81%       Indels:      11
DB:                12           Gaps:         3
US-09-997-610-2_COPY_18_149 (1-132) x BE757275 (1-504)
OY 2 GIYPROPRO-----AlahisProaRgProProGIUGlUaIGlYProProGIaYala 18
DB 181 GGAACCTCGGGACCTAGTGGCTTCTCCAGGCGCACCGTTATACAGGACCATAGGTARG 240
OY 19 ProGIaYleU-----ProGIaYrThrGIaYGIaYIleSerGIaYmetThrIyScYsPro 35
DB 241 CCAGAGACTAACAGGAGACCTGGCTCTCTGACACTGCTTGAGAAC-----285
OY 36 CysProaSPilEglarGserAlaPheThrValIySleuSerGIaYIySleuProleuPro 55
DB 286 TGCCACACCCCTGCTGCTTCTTCCGTCAGAGCTAGTGGCCCTTCCACAGACCC 345
OY 56 PheIySProlIeIlePheThrGIaYleuIyTrasnaIaGIaRgaSPleuIySGIaUa 75
```

```
DB 346 TCCACAGCCCATGCTCTTCCAGGAAGTTGTGTACACACCATCAGGCGCATTCGACCCCGCC 405
OY 76 MetGIaYValPheAlaCysaRgaYalProGIaYasrTYrTYrSerSerPheasPaYalGIuLeu 95
DB 406 ACTGTTGTGTTCAGTCAGCGACGCTCCCTGGTGTGTACACTTGGCTTGACATTTGAGATTG 465
OY 96 HlshIscYsIySValasnlIerIpleuMetarg 106
DB 466 TTTACAGATGCTGTCAAGTGGGTCTTAATGCGG 498
RESULT 4
BU068972          663 bp  mRNA  linear  EST 11-DEC-2001
LOCUS             BU068972
DEFINITION        BU068972 NIBB Mochii normalized Xenopus tailbud library Xenopus
ACCESSION         BU068972
VERSION           BU068972
KEYWORDS          EST.
SOURCE            African clawed frog.
  ORGANISM        Xenopus laevis
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
    Xenopodinae; Xenopus.
  1 (bases 1 to 663)
REFERENCE          Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-I,T. and Kohara
AUTHORS            Y.
                  Expressed genes in X. laevis embryo
TITLE              Unpublished (2001)
JOURNAL            Contact: Tadasu Shin-I
COMMENT            National Institute of Genetics
                  1111 Yata, Mishima, Shizuoka 411-8540, Japan
                  Tel: 81-559-81-6856
                  Fax: 81-559-81-6855
                  Email: tshin@genes.nig.ac.jp.
FEATURES
  source            Location/Qualifiers
                  1..663
                  /organism="Xenopus laevis"
                  /db_xref="taxon:8355"
                  /clone_lib="XL051f08"
                  /clone_lib="NIBB Mochii normalized Xenopus tailbud
                  library"
                  /tissue_type="whole embryo"
                  /dev_stage="stage 25"
                  /note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs
                  were oligo-dT primed and directionally cloned. Staging
                  according to Nieuwkoop and Faber. Library is substracted
                  and was constructed by N. Garrett and A.M. Zorn,
                  (Wellcome/CRC Institute)."
BASE COUNT         179 a 160 c 164 g 160 t
ORIGIN
Alignment Scores:
Pred. No.:         1,61e-09      Length:      663
Score:             181.00        Matches:     46
Percent Similarity: 42.55%      Conservative: 14
Best Local Similarity: 32.62%   Mismatches:  47
Query Match:       25.42%       Indels:      34
DB:                13           Gaps:         6
US-09-997-610-2_COPY_18_149 (1-132) x BU068972 (1-663)
OY 1 AlagIyProProAlahisProaRgProProGIUGlUaIGlYProProGIaYalaProGIy 20
DB 106 TCAGGGCCACCTGCTGTATGCAAGACCTCCA-----GACACACAGTGAACCTGGC 156
OY 21 LeuPro-----GlnTYrThrGIaYIleSer 29
DB 157 CGTCCGCAAGAAATCGACCTACGCCCCAGATGAGACAGTATTTGCCAAATATGGGA 216
OY 30 GluMetThrIyScYsProCysProaSP-----38
```

```
Db 217 GCCATTGATGCAATAAAACCTCATGACTATGCAGCAAAAAAGAAATGAGCA 276
Qy 39 -----ileguarSerAlapherThVallysleuSerGlylysleuProleuProPhe 56
Db 277 AACATTACGAATAATGCCAGCATTCAGTCACTGACACACGCCCTTCCTTACTGGCT 336
Qy 57 LysProIleIlePheThnglyValleuTyAsnAlaGlnArgAspleuLysGluAlaMet 76
Db 337 GAACCTAATAAGTTTGGAAACGTGTATACACGCAAGATGGCTACATCCACTTACT 396
Qy 77 GYValIlePheAlaCysArgValProGlyAsnTyPheSerPheAspValGluLeuHis 96
Db 397 GCGCGTTCAGCTGTGAATATCCAGCAATTATTACTTTCTTACCATGTC----- 447
Qy 97 HisCysLys---ValAsnIleTrpLeuMetArgLysGlnIleLeuAlaAsnLysGluL 115
Db 448 CATTTGTAAGTGGCAATGTTTGATGAT-----GCGTTGTCAAGAAATATGACCA 498
Qy 116 ile 116
Db 499 GTA 501

RESULT 5
CNS0540B 1092 bp DNA linear GSS 26-JUL-2000
LOCUS Tetraodon nigroviridis genome survey sequence T3 end of clone
011D22 of library A from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL320996.1 GI:9553880
VERSION ALS20996
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 1092)
AUTHORS Roest Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
ESTimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)

TITLE
JOURNAL
MEDLINE 20296633
PUBMED 10835645
REFERENCE 2 (bases 1 to 1092)
AUTHORS Crolius,H.R., Jallion,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C.,
Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)

TITLE
JOURNAL
MEDLINE 20359837
PUBMED 10899143
REFERENCE 3 (bases 1 to 1092)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000)
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES
source
1. 1092
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone_lib="A"
/clone_1lb="A"
/note="Genoscope sequence ID : CONA011DB1A2-end : T3"

BASE COUNT 224 a 301 c 308 g 245 t 14 others

ORIGIN
```

```
Alignment Scores:
Pred. No.: 1,59e-07 Length: 1092
Score: 165,50 Matches: 42
Best Local Similarity: 45,80% Conservative: 18
Best Local Similarity: 32,06% Mismatches: 52
Query Match: 23,24% Indels: 19
DB: 17 Gaps: 4

US-09-997-610-2_copy_18_149 (1-132) x CNS0540B (1-1092)
Qy 2 GYProProAlaHisProArgProProGluGluValGlyProProGlyValA----- 18
Db 528 GGGCCCTTACCGCCGACCGCTGGAGACCTGGTGTGATGTGAACCTGGTCTGTCT 587
Qy 19 ProGlyLeuProGlnTyTrhnglyGluIle-----Ser 29
Db 588 CCCGACCGCCTGGTCCCTCCCGGGAAGTTGTGTGAGAAGGAAGTGGGTGGTCA 647
Qy 30 GluMetTrhLysCysProCysProAspIleGluArgSerAlapherThVallysleuSer 49
Db 648 GTTATGTCAATATCCCTC-----ATGTCCTCTCAGCTGCTGTGACC 692
Qy 50 GlyLysLeuProLeuProPheLysProIleIlePheThnglyValleuTyAsnAlaGln 69
Db 693 ACACCTTACCCAGCTGCTGGCAGCCCTATTAACTTGACCAATTGTATACATGCTGAG 752
Qy 70 ArgAspleuLysGluAlaMetGlyValPheAlaCysArgValProGlyAsnTyPheSer 89
Db 753 AATCAGTATTAACCCCGAGTGTGGCATCTTACCTGCCAGGTCTCGAGTCACTTCTT 812
Qy 90 SerPheAspValGluLeuHis-----HisCysLysValAsnIleTrpLeuMetArgLys 107
Db 813 TCCTACAGCATTCATGTCATGTGCTCATGACCTGTGGTCTGTACAAAATGGCCAG 872
Qy 108 GlnIleuAlaAsnLysGluGluIleSerLys 118
Db 873 CCTATTATGTTCACTTATGACGAGTACACAG 905

RESULT 6
B494317 512 bp mRNA linear EST 08-AUG-2002
LOCUS B494317 MF01FSA cDNA Oryzias latipes cDNA clone MF01FSA024E16 5',
DEFINITION mRNA sequence.
ACCESSION B494317 GI:22146243
VERSION B494317
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 512)
AUTHORS Kohara,Y., Shin-I,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL unpublished (2001)
COMMENT Contact: Tadasu Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.

FEATURES
source
1. 512
/organism="Oryzias latipes"
/strain="vJ-R"
/db_xref="taxon:8090"
/clone="MF01FSA024E16"
/clone_1lb="MF01FSA cDNA"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="fry stage 40"
```



1-7-72, Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp/  
URL: <http://genome.gsc.riken.go.jp/>  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. *Genome Res.* 10 (10), 1611-1630 (2000)  
Wajik, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Wachiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura  
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. *Genome Res.*  
10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
Y., and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. *Genome Res.* 11 (2), 261-269 (2001)  
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, T., Aizawa  
K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and  
Hayashizaki, Y.  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences *Mamm. Genome* 12, 673-677 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
further details.  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

## FEATURES

### source

## Source

ORGIN	BASE COUNT	169 a	189 c	182 g	147 t	2 others
Alignment Scores:						
Pred. No.:	6.06e-07					Length: 689
Score:	157.50					Matches: 40
Percent Similarity:	43.20%					Conservative: 14
Best Local Similarity:	32.00%					Mismatches: 46
Query Match:	22.12%					Indels: 25
OB:	10					Gaps: 4

[illegible]

```

gene
CDS

1..2462
/gene="Col8a1"
200..2434
/gene="Col8a1"
/note="data source:MGI, source key:MGI:88463,
evidence:ISS
procollagen, type VIII, alpha 1
putative"
/codon_start=1
/protein_id="BAB31383.1"
/db_xref="GI:12858613"
/db_xref="MGI:MGI:88463"
/translation="MAVPPRPDLGLGLITISLNSVALIAGAYTGIKPLPQIIPQI
PQIQYQPLQGVQPHMPLKDKGLSMKCEKPHMQCYEYHLQIPKMEIRPPVPMKGE
VLPKKGCEVPLASLRNGCGPRGPRGPRGPGPLGHEMHPGLKYMKEIPQGVGIGK
PGMHGMPKPGAMGMPGAKGEIKGEIKGEMGILPGPGSPGPHLPLDGLPGGGLGPG
QPGKGGRRGPKGPGPGPLGGPKGEGKFGMDPLGLKPGPMHGPGLVGLPGVKGPG
VTGPRGQGLGKRPGEDEPGPOGLIVPGVQGFPGMAGYGKRGQDSIPQGRFPGK
GEGDLPLGPPGLPGVGKPGFPGKGDRIIGVPGVLGRGEGKPIGARGMGPGPE
PGLGLPGMPPGALIGFPGPKGEGVAGQGGPGGEGGLGFGPKPGLFLEVPG
GMRLLPGPIGPKGEGGKHLGLPLGVGLGLKPEPIIPEDQGLGGPGILGATYPSG
PIGPGIIGBKGEGGLGPPGPGFVGKPGVAGLHGPGKGAALGQGGPLGPPGPP
GPPGPPAVMPPTPGSEYELPDMGLGIDGVCPHAYAKKKGHGKATEMAFATLTV
PFPVPGAVKFDKLLYNGRONNPQTSITFCLEVGVITYAIVHNGKGNVALEKNN
EPMWYITDEKKGFLDQASGAVILLRPGDVFQMPSEDAAGLYACQVYHSSFGYL

```

[illegible]

QY	100	ValAsnIleTrrPleu	104
	:    :		
Db	2213	GGCAACGATCGGTT	2227
RESULT	10		
LOCUS	BQ419957		
DEFINITION	BQ419957	662 bp	mRNA
	fail2c05 y1 zebrafish fin day3 regeneration Danio rerio cDNA clone		linear
	5009744 5' similar to TR:Q9ZIK4 Q9ZIK4 COLLAGEN ALPHA 1 TYPE X		EST 23-MAY-2002
	;contains element MSRI repetitive element ; , mRNA sequence.		

ACCESSION B0419957  
 VERSION B0419957.1 GI:21125158  
 KEYWORDS EST.  
 SOURCE zebrafish.  
 ORGANISM Danio rerio.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.  
 1 (bases 1 to 662)  
 Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Peterson, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.  
 Washu Zebrafish EST Project 1998  
 Unpublished (1998)  
 Contact: Stephen L. Johnson  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: zbrafish@watsn.wustl.edu  
 CDNA Library Preparation: Raymond Lee. CDNA Library Arrayed by: Matthew Clark. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: Genome Systems, St. Louis, Missouri (web address: www.genomesystems.com) (email contact: info@genomesystems.com) and Research Genetics, Huntsville, Alabama (web address: www.resgen.com) (email contact: info@resgen.com) and Ressourcenzentrum Primatendatenbank, Berlin, Germany (web address: www.rzpd.de)  
 Seq primer: T3 ET from Amersham  
 High quality sequence stop: 468.  
 Location/Qualifiers  
 1..662  
 /organism="Danio rerio"  
 /db\_xref="taxon:7955"  
 /clone="590974"  
 /clone\_id="zebrafish fin day3 regeneration"  
 /sex="mixed male and female"  
 /tissue\_type="3 day fin regenerates"  
 /lab\_host="E. coli XL0LR"  
 /note="Vector: pBK-CMV. Site 1: EcoRI. Site 2: XhoI. 1st strand cDNA primed with (GA)10CTAGCTCTGAG(T18, followed by second strand synthesis, and ligated to 5' adapter (5'-)aatcgacagag-3', 3'-ggcgagcgc-5'. cDNA was cloned directionally (EcoRI/XhoI) into Stratagene Zap express lambda phage arms. Mass invivo excision done to obtain inserts in pBK-CMV phagemid."  
 BASE COUNT 161 a 192 c 152 g 157 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 6.48e-07 Length: 662  
 Score: 157.00 Matches: 43  
 Percent Similarity: 43.45% Conservative: 20  
 Best Local Similarity: 29.66% Mismatches: 46  
 Query Match: 22.05% Indels: 36  
 DB: 14 Gaps: 5  
 US-09-997-610-2\_COPY\_18\_149 (1-132) x B0419957 (1-662)  
 QY 2 glyProProAlaHisProArgPro-----ProGluGluValGlyPro 15  
 Db 210 GGGCCGCGGATGATCCGCGTGCCAGAGTCAGATGCTTCACAGTCACAGAGACT 269  
 QY 16 ProGlyAlaProGlyLeuProGlnTyrThrGlyGluIle----- 28  
 Db 270 CCGGACGACCTGGCCCTCCCA-----GGTCAGATTGTTTACACACGAGAGAGAGC 320  
 QY 29 -----serGluMetThrLysCysPro 35  
 Db 321 ATGCCAATTAAATCCACGAGATGGTGATGCTTTAGCCATGAGCTGATGAGAGGACACC 380

QY 36 CysProAspIleGluArgSerAlaPheThrValLysLeuSerGlyLysLeuProLeuPro 55  
 Db 381 -----ATGTCAGCATTCACAGCTCTTCACACAGCTTATCCTAATGCC 425  
 QY 56 PheLysProIleIlePheThrGlyValLeuTyrAsnAlaGlnArgAspLeuLysGluAla 75  
 Db 426 GGAACACCGATCGTGTTCACACCAAGATCGTACATGGAGAAACCACTATGTCACAC 485  
 QY 76 MetGlyValPheAlaCysArgValProGlyAsnTyrTyrSerSerPheAspValGluLeu 95  
 Db 486 AGGGCATATTCACCGCCAGGTCACAGACGTATTACTCTCTTCCTCCATAGCATGTC 545  
 QY 96 His-----HisCysLysValAsnIleTrrPleuMetArgLysGlnIleLeuAlaAsnLys 113  
 Db 546 AATGAGCAAAACCGCTTGTCACACTTACAAAACAAAGAGCCAGTCCTTTCTTAC 605  
 QY 114 GluGluIleSerLys 118  
 Db 606 GATGAGTACAAACAG 620  
 RESULT 11  
 AV243093  
 LOCUS 695 bp, mRNA linear, EST 24-OCT-2001  
 DEFINITION AV243093 RIKEN full-length enriched, 0 day neonate head mus  
 musculus cDNA clone 4831416g24 3, mRNA sequence.  
 AV243093  
 ACCESSION AV243093.2 GI:16386815  
 VERSION AV243093  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 695)  
 Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M., and Hayashizaki, Y.  
 RIKEN Mouse ESTs (Arakawa, T., et al. 2001)  
 Unpublished (2001)  
 On Nov 4, 1999 this sequence version replaced gi:6230520.  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@sc.riken.go.jp/  
 URL: http://genome-gsc.riken.go.jp/  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 Wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanka, I., Alzawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., and Hayashizaki, Y.



```

QY 14 GYProProglAlaProGlyLeuProGlnTyrThrGlyGluLeuSerGluMetThLys 33
Db 307 GGGCCACGGGGCTTCCCGAACCCTGGCAGAAAGAGAGAGCCGTGAGAACCGCTTAT 366
QY 34 CyProCysProAspIleGluArgSerAlaPheThrValLysLeuSerGlyLysLeuPro 53
Db 367 -----GIGTATCCGCTCAGCGCTTCAAGTGGGGCTGGAGACCCCGCTACT 411
QY 54 LeuProPheLysProIleIlePheThrGlyValLeuTyrAsnAlaGlnArgAspLeuLys 73
Db 412 GTTCCCAATGATACCATTCCTTACTACATCTTCTACACCAACAGCAATCATTTATGAC 471
QY 74 GluAlaMetGlyValPheAlaCysArgValProGlyAsnTyrTyrSerPheAspVal 93
Db 472 GGCAGCACTGGCAAGTTCTACTGCAACATTCGCGACTCTACTACTCTCTTACACATC 531
QY 94 GluLeuHis-----HisCysLysValAsnIleTyrLeuMetArgLysGlnIleLeuAla 111
Db 532 ACGGTGTACATGAAAGATGTGAAGGTGAGCCCTCTTCAAGAGACAGCCGCTTCTTTC 591
QY 112 AsnLysGluGluIleSerLysGlnGlnSerIleGlnGluValThrTyrValLeuLeu 130
Db 592 ACCTAGACCAATGATCAGAAAGAAATGTGGACCAAGCCCTCTGCATCTGTGCTCTTC 648

RESULT 13
LOCUS AW988554 468 bp mRNA linear EST 02-JUN-2000
DEFINITION ug08h07.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
IMAGE:1531069.5' similar to gb:U37222 Mus musculus 30kba adipocyte
complement-related protein Acrp30 (MOUSE);, mRNA sequence.
ACCESSION AM988554
VERSION AM988554.1 GI:8183731
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 468)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabps-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
MGI Consortium (info@image.jnl.gov) for further information.
MGI:947169
Seg primer: -40RP from Gibco
High quality sequence stop: 457.
FEATURES
source
1.468
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1531069"
/clone_1ib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker: 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT73 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 114 a 117 c 131 g 106 t
ORIGIN
Alignment Scores: 1.21e-06 Length: 468
Pred. No.: 152.50 Matches: 39
Score:

```

```

Percent Similarity: 49.58% Conservative: 20
Best Local Similarity: 32.77% Mismatches: 53
Query Match: 21.42% Indels: 7
DB: 10 Gaps: 2

US-09-997-610-2_COPY_18_149 (1-132) x AW988554 (1-468)
QY 14 GYProProglAlaProGlyLeuProGlnTyrThrGlyGluLeuSerGluMetThLys 33
Db 61 GGGCCACGGGGCTTCCCGAACCCTGGCAGAAAGAGAGAGCCGTGAGAACCGCTTAT 120
QY 34 CyProCysProAspIleGluArgSerAlaPheThrValLysLeuSerGlyLysLeuPro 53
Db 121 -----GIGTATCCGCTCAGCGCTTCAAGTGGGGCTGGAGACCCCGCTACT 165
QY 54 LeuProPheLysProIleIlePheThrGlyValLeuTyrAsnAlaGlnArgAspLeuLys 73
Db 166 GTTCCCAATGATACCATTCCTTACTACATCTTCTACACCAACAGCAATCATTTATGAC 225
QY 74 GluAlaMetGlyValPheAlaCysArgValProGlyAsnTyrTyrSerPheAspVal 93
Db 226 GGCAGCACTGGCAAGTTCTACTGCAACATTCGCGACTCTACTACTCTCTTACACATC 285
QY 94 GluLeuHis-----HisCysLysValAsnIleTyrLeuMetArgLysGlnIleLeuAla 111
Db 286 ACGGTGTACATGAAAGATGTGAAGGTGAGCCCTCTTCAAGAGACAGCCGCTTCTTTC 345
QY 112 AsnLysGluGluIleSerLysGlnGlnSerIleGlnGluValThrTyrValLeuLeu 130
Db 346 ACCTAGACCAATGATCAGAAAGAAATGTGGACCAAGCCCTCTGCATCTGTGCTCTTC 402

RESULT 14
LOCUS BE625509 552 bp mRNA linear EST 24-AUG-2000
DEFINITION uul9e06.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
IMAGE:3372418.5' similar to gb:U37222 Mus musculus 30kba adipocyte
complement-related protein Acrp30 (MOUSE);, mRNA sequence.
ACCESSION BE625509
VERSION BE625509.1 GI:9905925
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 552)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabps-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
MGI Consortium (info@image.jnl.gov) for further information.
MGI:1082022
Seg primer: -40RP from Gibco
High quality sequence stop: 486.
FEATURES
source
1.552
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3372418"
/clone_1ib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker: 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT73 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."

```

BASE COUNT	143 a	148 c	135 g	126 t
ALIGNMENT SCORES:				
Prod. NO.:	1.53e-06			552
Score:	152.50			39
Percent Similarity:	49.58%			20
Best local Similarity:	32.77%			53
Query Match:	21.42%			7
DB:	10			2
	Gaps:			
US-09-997-610-2_COPY_18_149 (1-132) x BE625509 (1-552)				
QY 14	GLYPRPGLGIALPAPROGLYLEUPROGLINTYRTHRGYLIIIESERGLUMETHRLYS 33			
DB 22	GGGCCACGGGGCTTCCCGGAACCCCTGGCAGAAAGAGAGACCTGGAGAACCCGCTTAT 81			
QY 34	CYSPPCYSPASPIELGLIARSERALAPHETHRLVALLYSLEUSERGLYLSLEUPRO 53			
DB 82	-----GTTATGCCCTCACGCTTCACTGTGGGCTGTGAGACCCGCTCACT 126			
QY 54	LEUPROPHIELYSPROIIELEPHERTHRGYLVALEUTYRASNALGLNARGASPLEULYS 73			
DB 127	GTTCCTCAATGTACCATTCGTCTTACTAAGACTCTTCTCAACACACCAATCATTTATGAC 186			
QY 74	GLIUALMECTGLYALPHEALACYSARGYALPROGLYASNTYRTYTSERSEHPASVAL 93			
DB 187	GGCAGACACTGGCACTTACTGTACGAAACATTCGGGACCTCTACTCTTCTTACACATC 246			
QY 94	GLUTEUHS-----HISCYSLYSVALASNILETRPLEUMETARGYSLINLEUALA 111			
DB 247	ACGGGTGCATGAAAGATGTGAAGGTGACCCCTTTCACAGAAAGACAGGCCCTTCTTC 306			
QY 112	ASNLYSGUGLIULESERLYSGLINSEIRIEGLINGLIUALTHRTYRVALLEULEU 130			
DB 307	ACCTACGACACGATACAGAAAGAATGTGGACGACGCTGTGGCTGTGCTCCTC 363			
RESULT 15				
LOCUS	BB625284	633 bp	mRNA	linear
DEFINITION	BB625284 RIKEN full-length enriched, adult male cecum Mus musculus			
ACCESSION	BB625284	CDNA clone	9130217622 5', mRNA sequence.	
VERSION	BB625284.1	GI:16463576		
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 633) Arakawa,T., Carrinck,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Kono,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,K., Sano,H., Sasak,i., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.			
TITLE	RIKEN Mouse ESTs (Arakawa,T., et al. 2001)			
JOURNAL	Unpublished (2001)			
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/ Carrinck,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)			

**FEATURES**  
**SOURCE**

Waki K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Matsubara, M., Itonaka, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga,  
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and  
Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10  
(11), 1757-1771 (2000)  
Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carrinci, P., Sugahara,  
Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)  
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa,  
K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and  
Hayashizaki, Y.  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences. *Mamm. Genome* 12, 673-677 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.  
e mouse tissues. *Genome*

BASE COUNT	168 a	152 c	194 g	119 t
ORIGIN				
Alignment Scores:				
Prod. No.:	1.87e-06	Length:	633	
Score:	152.50	Matches:	42	
Percent Similarity:	46.77%	Conservative:	16	
Best local Similarity:	33.87%	Mismatches:	55	
Query Match:	21.42%	Indels:	11	
DB:	10	Gaps:	4	
US-09-997-610-2_COPY_18_149 (1-132) x BB625284 (1-633)				
Oy	2	gLyPProAlaHisProArgProProGluLuAluAlGlyPro-----	15	
Db	202	GGCCCCCATCGCCAAACCAACGATCCACAGGGGAAGACTGTGACCCATAGGGCCCCCAGGGGAG	261	
Oy	16	ProGluAlaProGlyLeuProGluIntYrthGlyLuIleSerGluMetYrthLysCys---	34	
Db	262	CCAGAGATCAAGAGGAATGACAGCTGCAGAAAGCATCGAGAGAGAGGGAAGATTGCT	321	
Oy	35	ProCysProAlaPheGluArgSerAlaPheYrthValLysLeuSer-----GlyLysLeu	52	
Db	322	GAGCTCCCTTGAGCCCAAGATGACTTCACTGTGGAGCACTACAGGTGATCAGTAAAGTTTC	381	

```
QY 53 ProLeuProPheLysProIleIlePheThrGlyValLeuTyrAsnAlaGlnArgAspLeu 72
    ||| ||| ||||| ||| :||| |||||
Db 382 CCTCCCCAGATGCACCCATTAAATTGCAATGATCCTATACATGAAGTGAACCACTAC 441
QY 73 LysGlnAlaMetGlyValPheAlaCysArgValProGlyAsnTyrTyrSerSerPheAsp 92
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 442 AATGTAGCGAGGGAAGTTCCACCTGCCACGTGCGACAGTGTCTATATTACTTTACCTACCAT 501
QY 93 ValGluLeuHisHisCysLysValAsnIleTrpLeuMetArgLys-----GlnIleLeu 110
    ::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 502 ATCACTGTGTTCCTCAAGAATGTCAGGTATCTTTGGTCAAAAACGGGGTAAAGTCTCTG 561
QY 111 AlaAsnLysGlu 114
    |||||
Db 562 CACACCAAGAA 573
```

Search completed: February 19, 2003, 19:35:51  
Job time : 1032.49 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 19, 2003, 14:57:18 ; Search time 1839.21 Seconds  
(without alignments)  
2088.706 Million cell updates/sec

Title: US-09-997-610-2\_COPY\_18\_149  
Perfect score: 712  
Sequence: 1 AGPRAHRRPREVGPARGA.....KEELKQGSIDEVTVLLKA 132

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_p2n.model -DEV=x1h  
-O=/cgn2.1/USPTO-SPool/US09997610/runac\_10022003.160823.23819/app\_query.fasta\_1.1635  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIG=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US99997610.ecgn.1.1.3965.etunac\_10022003.160823.23819 -NCPu=6 -ICPU=3  
-NO\_XLIPY -NO\_MAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: GenEmbl:\*  
2: gb\_da:\*  
3: gb\_htg:\*  
4: gb\_in:\*  
5: gb\_om:\*  
6: gb\_ov:\*  
7: gb\_pat:\*  
8: gb\_ph:\*  
9: gb\_pl:\*  
10: gb\_pr:\*  
11: gb\_ro:\*  
12: gb\_sts:\*  
13: gb\_sy:\*  
14: gb\_un:\*  
15: gb\_vl:\*  
16: em\_da:\*  
17: em\_fun:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_ph:\*  
24: em\_pi:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pin:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vit:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	623	87.5	145880	9 HS302D9	282198 Human DNA s
2	320.5	45.0	1265	10 TMSHP20A	D12974 Tamias asia
3	290	40.7	5121	10 AB067779	AB067779 Tamias si
4	217	30.5	2005	10 TMSHP25	D12975 Tamias asia
5	169.5	23.8	1918	10 RNO131848	AJ131848 Rattus no
6	167.5	23.5	2609	5 CHKX	M13496 Chicken typ
7	167	23.5	559	10 AB06781363	AB06781363 Callosciu
8	167	23.5	1385	10 TMSHP27	D12976 Tamias asia
9	163.5	23.0	2235	6 AX332258	AX332258 Sequence
10	163.5	23.0	2235	9 HSCOL8A1	X57577 Human COL8A
11	163.5	23.0	2415	4 RABCOLV111	J05042 Rabbit type
12	162.5	22.8	2506	9 BC013581	BC013581 Homo sapi
13	162.5	22.8	117000	9 AC069222	AC069222 Homo sapi
14	162.5	22.8	145880	9 HS302D9	282198 Human DNA s
15	162.5	22.8	152354	2 AC120598	AC120598 Homo sapi
16	162.5	22.8	169922	2 AC022883	AC022883 Homo sapi
17	162.5	22.8	182978	30 AC067824	AC067824 Homo sapi
18	161	22.6	7089	4 AF222861	AF222861 Sus scrofa
19	159.5	22.4	1986	10 MMCOL8A	X66977 M.musculus
20	159	22.3	3143	4 BTCOL10A1	X53556 Bovine COL1
21	158	22.2	3422	10 MNA1TXCOL	Z21610 M.musculus
22	158	22.2	9331	10 MNCOL10A	X67348 Mus musculus
23	158	22.2	68834	2 AC119229	AC119229 Mus muscu
24	158	22.2	179303	2 AC016087	AC016087 Homo sapi
25	158	22.2	206213	2 AC021709	AC021709 Mus muscu
26	157.5	22.1	2435	10 BC011061	BC011061 Mus muscu
27	156	21.9	731	10 MNA1XCOL	X63013 M.musculus
28	156	21.9	1973	9 HS10A1COL	X65120 H.sapiens C
29	156	21.9	3215	9 HSCOLX3	X72580 Homo sapien
30	156	21.9	3226	6 AX333243	AX333243 Sequence
31	156	21.9	3226	9 HSCOLAIIX	X60382 H.sapiens C
32	156	21.9	3226	11 G28608	G28608 human STS S
33	156	21.9	10058	9 HSCOLIX	X98568 H.sapiens t
34	156	21.9	107553	9 HSJ136014	AL121963 Human DNA
35	156	21.9	205594	2 AL355373	AL355373 Homo sapi
36	155.5	21.8	582	4 AF417206	AF417206 Canis fam
37	152.5	21.4	947	10 BC028770	BC028770 Bos tauru
38	152.5	21.4	1134	4 AF269230	AF269230 Bos tauru
39	152.5	21.4	200942	2 AC125396	AC125396 Mus muscu
40	150.5	21.1	1152	6 AX195207	AX195207 Sequence
41	150.5	21.1	1152	6 AX358517	AX358517 Sequence
42	150.5	21.1	1152	6 MM049915	MM049915 Sequence
43	150.5	21.1	1276	6 AR034252	AR034252 Sequence
44	150.5	21.1	1276	6 AX195209	AX195209 Sequence
45	150.5	21.1	1276	6 AX358519	AX358519 Sequence

RESULT 1

## ALIGNMENTS

[illegible]

```

misc_feature      /note="match: GSS: Em:AQ553482"
14616..15060
/note="match: GSS: Em:AQ370601"
repeat_region     14868..15040
/note="MIR repeat: matches 49..233 of consensus"
repeat_region     15071..15188
15071..15188
/note="L2 repeat: matches 2112..2239 of consensus"
repeat_region     15304..15399
15304..15399
/note="MUT1B repeat: matches 1..99 of consensus"
repeat_region     15490..15662
15490..15662
/note="ALUSg1 repeat: matches 2..114 of consensus"
repeat_region     15669..15727
15669..15727
/note="MUT1B repeat: matches 119..178 of consensus"
repeat_region     15728..16027
15728..16027
/note="ALUSc repeat: matches 1..299 of consensus"
repeat_region     16028..16245
16028..16245
/note="MUT1B repeat: matches 178..390 of consensus"
repeat_region     16346..16834
16346..16834
/note="AluY repeat: matches 1..300 of consensus"
repeat_region     18296..18323
18296..18323
/note="MSTA repeat: matches 2..29 of consensus"
repeat_region     18324..18392
18324..18392
/note="MER66-internal repeat: matches 4919..4993 of consensus"
repeat_region     18393..18712
18393..18712
/note="AluDb repeat: matches 1..311 of consensus"
repeat_region     18713..19133
18713..19133
/note="MER66-internal repeat: matches 4548..4919 of consensus"
misc_feature      complement(18872..19230)
/note="match: GSS: Em:AQ05063"
19251..19719
/note="match: GSS: Em:BI4179"
19537..20230
/note="HERVFN21 repeat: matches 4657..5784 of consensus"
repeat_region     20317..20382
20317..20382
/note="33 copies 2 mer ta 68 conserved"
repeat_region     20513..20666
20513..20666
/note="77 copies 2 mer tt 70 conserved"
repeat_region     20682..21008
20682..21008
/note="AluSg1 repeat: matches 1..306 of consensus"
repeat_region     21239..21533
21239..21533
/note="HMER-P3 repeat: matches 4410..4713 of consensus"
repeat_region     21882..22254
21882..22254
/note="THE1B repeat: matches 1..364 of consensus"
repeat_region     22302..22337
22302..22337
/note="MER66-internal repeat: matches 2186..2417 of consensus"
repeat_region     22538..22650
22538..22650
/note="AluSp repeat: matches 1..313 of consensus"
repeat_region     22851..23801
22851..23801
/note="MER66-internal repeat: matches 1210..2186 of consensus"
repeat_region     23905..23989
23905..23989
/note="MER66-internal repeat: matches 3017..3102 of consensus"

Alignment Scores:
Pred. No.:      6.49e-47      Length:      145880
Score:          623.00      Matches:      117
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    87.50%      Indels:      0
DB:             9           Gaps:         0

US-09-997-610-2_COPY_18_149 (1-132) x HS302D9 (1-145880)
OY 16 ProGlyAlaProGlyLeuProGlyIleuProGlyGluIleSerGluMetThrIleCysPro 35
DB 37262 CCTGGGCCCCAGGTTTACACACATATACAGGAGAAATAGCAAAATGCCCC 37321
OY 36 CysProAspIleGluArgSerAlaPheThrValIleSerGlyLeuProLeuPro 55
DB 37322 TGTCCGATATAGTAAGGACGCTTACTGTGAAGCTCAGTGAAGAACTCTCTTCT 37381

```

```

OY 56 PhelysProIleIlePheThrGlyValLeuTyAsnAlaGlnArgAspLeuTysGluAla 75
DB 37382 TTCAGCCCATCATCTTTCACAGGGGTCCTGTACAAATGCCAGAGGGATTAAAGAGGCC 37441
OY 76 MetGlyValPheAlaCysArgValProGlyAsnTyrTyrSerSerPheAspValGluLeu 95
DB 37442 ATGGAGTCTTTCCTTCAGAGGTCCTGGGAATTAAGTCTTTCAGTGTGAAGCTG 37501
OY 96 HisHisCysIleValAsnIleIlePheMetArgIleGlnIleLeuAlaAsnIleGlu 115
DB 37502 CATCATTCGACAGGTGATATTTGGCTATGAGAACCAATTTTGGCTAATTAAGAGAA 37561
OY 116 IleSerIleGlnIleSerIleGlnIleGluValThrTyrValLeuLeuTysAla 132
DB 37562 ATTCTAAGCAGCAAGCATTCAGAGGTCCTGGTGTCTGTAAAGCA 37612

RESULT 2
TMSHP20A      1265 bp      mRNA      linear      ROD 03-FEB-1999
LOCUS         TMSHP20A
DEFINITION    Tamias asiaticus mRNA for HP-20, complete cds.
ACCESSION     D12974
VERSION       D12974.1 GI:287467
KEYWORDS      HP-20; collagen-like domain; hibernation-related protein; plasma protein.
SOURCE        Tamias asiaticus liver cDNA to mRNA, clone:PCM20-7.
ORGANISM      Tamias sibiricus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Scluridae; Sciurinae;
Tamias.
REFERENCE     1 (bases 1 to 1265)
AUTHORS      Takamatsu N., Ohba K., Kondo J., Kondo N. and Shiba T.
TITLE        Hibernation-associated gene regulation of plasma proteins with a collagen-like domain in mammalian hibernators
JOURNAL       Mol. Cell. Biol. 13 (3), 1516-1521 (1993)
MEDLINE      93180798
REFERENCE     2 (bases 1 to 1265)
AUTHORS      Takamatsu N.
TITLE        Direct Submission
JOURNAL       Submitted (21-AUG-1992) Nobuhiko Takamatsu, School of Hygienic Sciences, Kitasato University, Laboratory of Molecular Biology; 1-15-1 Kitasato, Sagamihara, Kanagawa 228, Japan (Tel:0427-78-9408, Fax:0427-78-9403)

FEATURES
source        1..1265
              /organism="Tamias sibiricus"
              /db_xref="taxon:64680"
              /clone="PCM20-7"
              /tissue_type="liver"
              /note="collagen-like domain at nt 161-277"
              /codon_start=1
              /product="HP-20"
              /protein_id="BAA02351.1"
              /db_xref="GI:287468"
              /translation="MTDVRLAIFVLMAVNLNDQYSCSGPPRPVGRPGVGRPGP
              PGQPGAGRPGRPGKPSVKPCRCRRSAFTVKFGRLRPSRPVFEVLNTGRDL
              KESTGFNCVEPDEGNVHFSDVELVHCKVIGLKHIVMERKHOLSKMEYENASGAMI
              MRLROGDKVWLEADVETEBPDQAKVIYSGFLISS"
              89..157
              sig_peptide
              mat_peptide
              158..676
              /product="HP-20"
              /protein_id="HP-20"
              1238..1243
              polyA_site
              polyA_site
              1265
              BASE COUNT      354 a      304 c      288 g      319 t
ORIGIN
Alignment Scores:
Pred. No.:      6.29e-21      Length:      1265
Score:          320.50      Matches:      66
Percent Similarity: 65.00%      Conservative: 12
Best Local Similarity: 55.00%      Mismatches: 39
Query Match:    45.01%      Indels:      3

```

```

DB: 10 Gaps: 2
US-09-997-610-2_copy_18_149 (1-132) x TMSHP20A (1-1265)
QY 2 GlyProAlaHisProArgProGluGluValGlyProGlyAlaProGlyLeu 21
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 GGAGTGCCTGGAGATGCGGGGCTCGTGGCCCCCAGAGTCAACAGCGGGGAGG 247
QY 22 ProGlnIleThrGlyLeuIleSerGluMetThrLysScyProCysProAspIleGluArg 41
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 248 CCAGAGAGACCCAGACCAAAAGGCCCAAGTAAATAGCCCTGCAGAG-----GAGAGG 301
QY 42 SerAlaPheThrValLysLeuSerGlyLysLeuProLeuProPheLysProIleIlePhe 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 302 TCAGCCTTCACGGTGAAGTTCACAGCGGAAGGCTCCTCCACCTCGAGAGCTGTGCTTC 361
QY 62 ThrGlyValLeuThrValAsnAlaGlnArgAspLeuLysGluAlaMetGlyValAlaPheAlaCys 81
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 ACAGAGGTCTCTGACATACCCAGAGGAGCTTCAAGAGACAGAGAGCTTTAACTGC 421
QY 82 ArgValProGlyAsnThrLysSerSerPheAspValGluLeuHisScyLysValAsn 101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 422 GTGAGCCTCGAATATACCATTTACAGCTTGATGTTGAGCTTACCACTCAAGGTGAAG 481
QY 102 IleThrLeuMetLysGlnIle---LeuAlaAsnLysGluGluIleSerLysGlnGln 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 482 ATTTGGGTTGATGAAGACCAATCAAGTCATGCAAGAAAGCATCAGCTTCCAAAATGAA 541

RESULT 3
AB067779 5121 bp DNA linear ROD 26-OCT-2001
LOCUS Tamias sibiricus gene for HP-20, complete cds.
DEFINITION AB067779
VERSION AB067779.1 GI:15706341
KEYWORDS
SOURCE Tamias sibiricus DNA, clone:lamdacM2061.
ORGANISM Tamias sibiricus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Scuriidae; Sciurinae;
          Tamias.
REFERENCE
  1 Ono,M., Hosoe,Y., Azuma,S., Shoji,M., Nara,K., Kondo,N., Shiba,T.
    and Takamatsu,N.
    HNF-1 regulates the liver-specific transcription of the chipmunk
    HP-20 gene
    JOURNAL Gene 277 (1-2), 121-127 (2001)
    MEDLINE 21488336
    REFERENCE 2 (bases 1 to 5121)
    AUTHORS Ono,M.
    TITLE Direct Submission
    JOURNAL Submitted (04-AUG-2001) Motoharu Ono, Kitasato University,
    Department of Biosciences, School of Science, 1-15-1 Kitasato,
    Sagamihara, Kanagawa 228-8555, Japan
    (E-mail:ms9805mestu.sci.kitasato-u.ac.jp, Tel:81427789408,
    Fax:81427789408)

FEATURES
  source          Location/Qualifiers
    1..5121
    /organism="Tamias sibiricus"
    /db_xref="taxon:64680"
    /clone="lamdacM2061"
    /note="synonym:Tamias asiaticus"
    1518..1684
    join(1678..1684,2573..2710,3968..4413)
    /gene="HP-20"
    join(1678..1684,2573..2710,3968..4413)
    /gene="HP-20"
    /codon_start=1
    /product="HP-20"
    /protein_id="BAB68362.1"
    /db_xref="GI:15706342"
    /translation="MTDAMRLAIFVLMVNVLDNDYSCSGPPVGVGPGVPGRGP
    PGQPGAGRPDGPGRGSPVKPCPKERSFIVKESGRISLPSEPVVFTEVLNTQIDL
    KASTGVNCEPBGNYHFSFDVELYHCKYKIGIMKNHIOYMERKHOLSKNEYENASGAMI
  
```

```

exon          MFLRGDKYWLADVETEEDDAKKVIYFSGELLIS"
2573..2710
/gene="HP-20"
BASP COUNT 1571 a 1018 c 1090 g 1442 t
ORIGIN
Alignment Scores:
Pred. No.: 2,01e-17 Length: 5121
Score: 290.00 Matches: 60
Percent Similarity: 66.97% Conservative: 13
Best Local Similarity: 55.05% Mismatches: 30
Query Match: 40.73% Indels: 6
DB: 10 Gaps: 3

US-09-997-610-2_copy_18_149 (1-132) x AB067779 (1-5121)
QY 13 ValGlyProProGlyAlaProGlyLeuProGlnIleThrGlyLeuIleSerGluMetThr 32
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3964 CTAGGCGCTCGGGGTAGGACGAGACCA-----GGACCAAAAGGCCCAAGTGA 4014
QY 33 LysCysProCysProAspIleGluArgSerAlaPheThrValLysLeuSerGlyLysLeu 52
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4015 AAATGCCCTGCAGCA-----GAGAGCTCAGCCTTCACGCTGAAGTTCACGCGAAGCTC 4068
QY 53 ProLeuProPheLysProIleIlePheThrGlyValLeuThrValAsnAlaGlnArgAspLeu 72
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4069 CTTCCACCTTCGAGAGCCTGGTGTCTTCAAGAGGCTCTGACATACCCAGAGGAGCTTG 4128
QY 73 LysGluAlaMetGlyValAlaPheAlaCysArgValProGlyAsnThrLysSerSerPheAsp 92
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4129 AAGCGAGCACAGAGAGCTTTAACTGCGTGAGAGCTGGAATTAACATTCAGCTTTGAT 4188
QY 93 ValGluLeuHisScyLysValAsnIleThrLeuMetLysGlnIle---LeuAla 111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4189 GTTGAGCTTTACCACTCAAGGTGAAGATTGGTTGATGAAGACCAATCAAGTCATG 4248
QY 112 AsnLysGluGluIleSerLysGlnGln 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4249 GAAAAGCATCAGCTCTCCAAAACGAA 4275

RESULT 4
TMSHP25
LOCUS TMSHP25 2005 bp mRNA linear ROD 03-FEB-1999
DEFINITION Tamias asiaticus mRNA for HP-25, complete cds.
ACCESSION D12975
VERSION D12975.1 GI:287469
KEYWORDS HP-25; collagen-like domain; hibernation-related protein; plasma
protein.
SOURCE Tamias asiaticus liver cDNA to mRNA, clone:PCM25-3.
ORGANISM Tamias sibiricus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Scuriidae; Sciurinae;
          Tamias.
REFERENCE
  1 (bases 1 to 2005)
  Takamatsu,N., Ohba,K., Kondo,J., Kondo,N. and Shiba,T.
  Hibernation-associated gene regulation of plasma proteins with a
  collagen-like domain in mammalian hibernators
  JOURNAL Mol. Cell. Biol. 13 (3), 1516-1521 (1993)
  MEDLINE 93180798
  REFERENCE 2 (bases 1 to 2005)
  Takamatsu,N.
  Direct Submission
  JOURNAL Submitted (21-AUG-1992) Nobuhiko Takamatsu, School of Hygienic
  Sciences, Kitasato University, Laboratory of Molecular Biology,
  1-15-1 Kitasato, Sagamihara, Kanagawa 228, Japan (Tel:0427-78-9408,
  Fax:0427-78-9403)

FEATURES
  source          Location/Qualifiers
    1..2005
    /organism="Tamias sibiricus"
    /db_xref="taxon:64680"
    /clone="PCM25-3"
    /issue_type="liver"
  
```









REFERENCE	AUTHORS	TITLE
JOURNAL MEDLINE PUBMED	Yamauchi, N., Benay, P. D., van der Rest, M., and Ninomiya, Y.	The cloning and sequencing of alpha 1(VIII) collagen cDNAs demonstrate that type VIII collagen is a short chain collagen and contains triple-helical and carboxyl-terminal non-triple-helical domains similar to those of type X collagen
89380199	J. Biol. Chem. 264 (27), 16022-16029 (1989)	
2476437		
COMMENT	Draft entry and computer-readable copy of sequence [1] kindly submitted by Y. Ninomiya 23-AGC-89.	
FEATURES	Location/Qualifiers	
source	1. .2415	
	/organism="Oryctolagus cuniculus"	
	/db_xref="taxon:9986"	
CDS	181. .2415	
	/note="alpha-1 (VIII) collagen precursor"	
	/protein_start=1	
	/codon_id="AAA1204.1"	
	/db_xref="GI:164896"	
	/translation="MAVPPGQQLQVLTLSGSLTLDAGATYGRKPLPPQIPOM PPDLPQPLGQVPHMPLAKDLTWMKEMPHAYCKEYHPLQYMKVEVPPMKKE AVRKKEIPLASLRGEQPRGEGEPGPPGRLPGQIGLKGKPGQYDVGKP GMKPGKPGAMSGKAGEIGEGELGPMGIGPPGPPGPHGIGLKGKPGQPLPGQ PGKAGDGRGPPGPPGLOGPKGKGGMGLGLKGPMPMHGPPGPPGPPGPPGPPG TGGPPGKPGKPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG EQLPLGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG GLPLGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG IRLGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG IGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG PGKPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG PEPVPGPIKEDRLVNGRONYDQGTIFPCVPGYVYFAYVHNHCKGNVVALFKNN EPVMTYDEYKKGFLDQASSAVILLRPPGVFLQMPSEDAAGLYAGQIVHSSPGTL LYPD"	
	181. .240	
sig_peptide	/note="alpha-1 (VIII) collagen signal peptide"	
mat_peptide	241. .2412	
	/product="alpha-1 (VIII) collagen"	
BASE COUNT	536 a	719 c 768 g 392 t
ORIGIN		
Alignment Scores:		
Pred. NO.:	5.63e-06	Length: 2415
Score:	163.50	Matches: 45
Percent Similarity:	44.20%	Conservative: 16
Best Local Similarity:	32.61%	Mismatches: 48
Query Match:	22.96%	Indels: 29
DB:	4	Gaps: 7
US-09-997-610-2_COPY_18_149 (1-132) x RABCOLVIII (1-2415)		
QY 2	GI:ProProAlaNIhSPROATgProProGIuGIaValaGIProProGIaIaProGIuYeu 21	
Db 1840	GGCCAGCCTGCGCTCCAGCGGCCCCAGAGCCCTCCAGAGGCTCCAGAGGCTCGT 1899	
QY 22	-----ProGInTYrThGlyGIu---IleSerGIuMet----- 31	
Db 1900	ATGGCCCCGACACAGCAGCAGCCAGGAGAGATATCTGCCAGATATAGGGGCTGGAGATTGAT 1959	
QY 32	---ThrLysCysPro-----CysProAspIle 39	
Db 1960	GGGTAATAACTCTCACGCTATGCGGCCAGAAAGGCAAGAACGGCGGCGCCAGCTAT 2019	
QY 40	GIuArSerAlaPheThrValaLysLeuSerGIuLysLeuProLeuProPheLysProIle 59	
Db 2020	GAGATGCGCTGCTTACCGCGGAGCTAGCAGCACCTTCCCGCGGTGGGGCCCGCAT 2079	
QY 60	IlePheThrGIuValaLeuTYrAsnIaGIaIaTrGAspIleuLysGIuAlaIaMetGIuValaPhe 79	

[illegible]







```

repeat_region 15304..15399
/note="MLT18 repeat: matches 1..99 of consensus"
repeat_region 15490..15662
/note="Alusgl repeat: matches 2..114 of consensus"
repeat_region 15669..15727
/note="MLT18 repeat: matches 119..178 of consensus"
repeat_region 15728..16027
/note="AlusC repeat: matches 1..239 of consensus"
repeat_region 16028..16245
/note="MLT18 repeat: matches 178..390 of consensus"
repeat_region 16546..16854
/note="AluY repeat: matches 1..300 of consensus"
repeat_region 18296..18323
/note="MSTA repeat: matches 2..29 of consensus"
repeat_region 18324..18392
/note="MER66-Internal repeat: matches 4919..4993 of consensus"
repeat_region 18393..18712
/note="Aluub repeat: matches 1..311 of consensus"
repeat_region 18713..19133
/note="MER66-Internal repeat: matches 4548..4919 of consensus"
misc_feature complement(18872..19230)
/note="match: GSS: Em:AQ005063"
misc_feature 19251..19719
/note="match: GSS: Em:BI4179"
repeat_region 19537..20290
/note="HERVH21 repeat: matches 4657..5784 of consensus"
repeat_region 20317..20382
/note="33 copies 2 mer ta 68 conserved"
repeat_region 20513..20666
/note="77 copies 2 mer tt 70 conserved"
repeat_region 20682..21008
/note="Alusgl repeat: matches 1..306 of consensus"
repeat_region 21239..21553
/note="HUEK-P3 repeat: matches 4410..4713 of consensus"
repeat_region 21882..22254
/note="THE1B repeat: matches 1..364 of consensus"
repeat_region 22302..22537
/note="MER66-Internal repeat: matches 2186..2417 of consensus"
repeat_region 22538..22850
/note="AlusP repeat: matches 1..313 of consensus"
repeat_region 22851..23801
/note="MER66-Internal repeat: matches 1210..2186 of consensus"
repeat_region 23905..23989
/note="MER66-Internal repeat: matches 3017..3102 of consensus"

Alignment Scores:
Pred. No.: 0.000595 Length: 145880
Score: 162.50 Matches: 42
Percent Similarity: 50.898 Conservative: 15
Best Local Similarity: 37.508 Mismatches: 46
Query Match: 22.82% Indels: 9
DB: 9 Gaps: 3

US-09-997-610-2_COPY_18_149 (1-132) x HS302D9 (1-145880)
QY 20 GtyleuproglintyrthrlgylulieserglumethrlyscyspProcysProaspille 39
DB 90953 GGAATTCCTCGATTCTATGAGACA-----CAAGACCAAGACAGCAGTACAGAGCATATT 90900
QY 40 Gtlu-----ArgserAlapheThrValLysLysenSerGylsLeupro 53
DB 90899 GAAAAAGCTTCATCCCAACCAAGATCTGCATTGTGCTGAAGCTGATCGACCCCTCCCA 90840
QY 54 LeupropheLysProLlellepheThrGylValLeuTyraSnAlaGlnArgspLeuLys 73
DB 90839 GTGCCCTTCTAGCGCATCTGCTTCAAGAGAGCCCTTATATAATACAGTTCCTTCAT 90780
QY 74 GtluAlaMeClYValPheAlaLacySarGValProGlyAsnTyTyrSerSerPheaspVal 93

```

```

DB 90779 TTCTGTGGTGAAGGATTACCTCTGTACCCATCCCTAGTGTGTACGACCTTTGGTTGAATT 90720
QY 94 GtluEuhHisCysLysValAsnIleTrpLeuMetArgLysIleLeuAlaAsn 112
DB 90719 GACCTGTTTCAGAGCTGTGCAATGTGGTATTTATGAGATGCGATTCATGACAAAT 90660
QY 113 LysGlucLulIeSerLysGlnGlnSerIleGlnL 124
DB 90659 AAMCATGCTGTGCGCAAGATGCTGTGACAGCA 90624

RESULT 15
AC120598 152354 bp DNA linear HTG 23-JUL-2002
AC120598
LOCUS
DEFINITION
AC120598
AC120598.3 GI:21902634
VERSION
KEYWORDS
HTG: HTGS_PHASE1.
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 152354)
Muzny,D.M., Adams,C., Adio-Oduola,B., All-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayala,M., Banks,T.,
Barbarta,J., Benton,J., Bimaye,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Dublin,K.J.,
Einhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulik,S., Hume,T., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S.,
Karlsdon,E., Kelly,S., Khan,U., King,L., Kovah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozodo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., MaWhiney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okuwonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peterson,L., Pickens,R., Prins,E., Pul,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., RojudoKani,L., Rolfe,M., Rolfe,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,
Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Taneey,J., Taylor,C., Taylor,T., Telford,B., Thomas,R., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 152354)
Worley,K.C.
Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 152354)
Worley,K.C.
Direct Submission

```

## JOURNAL

## COMMENT

Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jul 18, 2002 this sequence version replaced gi:20564266.

## ----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

## ----- Project Information

Center project name: GWNU

## ----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 142151 bases at least Q40

Consensus quality: 147764 bases at least Q30

Consensus quality: 152093 bases at least Q20

-----

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 33 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

1 2152: contig of 2152 bp in length  
\* 2153 2252: gap of unknown length  
\* 2253 4902: contig of 2650 bp in length  
\* 4903 5002: gap of unknown length  
\* 5003 7448: contig of 2446 bp in length  
\* 7449 7549: gap of unknown length  
\* 7549 10685: contig of 3137 bp in length  
\* 10686 10785: gap of unknown length  
\* 10786 13353: contig of 2568 bp in length  
\* 13354 13453: gap of unknown length  
\* 13454 16709: contig of 3256 bp in length  
\* 16710 16809: gap of unknown length  
\* 16810 20018: contig of 3209 bp in length  
\* 20019 20118: gap of unknown length  
\* 20119 23425: contig of 3307 bp in length  
\* 23426 23525: gap of unknown length  
\* 23526 26280: contig of 2755 bp in length  
\* 26281 26380: gap of unknown length  
\* 26381 29423: contig of 3043 bp in length  
\* 29424 29523: gap of unknown length  
\* 29524 32163: contig of 2640 bp in length  
\* 32164 32263: gap of unknown length  
\* 32264 36076: contig of 3813 bp in length  
\* 36077 36176: gap of unknown length  
\* 36177 39483: contig of 3307 bp in length  
\* 39484 39583: gap of unknown length  
\* 39584 43134: contig of 3551 bp in length  
\* 43135 43234: gap of unknown length  
\* 43235 46834: contig of 3600 bp in length  
\* 46835 46934: gap of unknown length  
\* 46935 52040: contig of 5106 bp in length  
\* 52041 52140: gap of unknown length  
\* 52141 57672: contig of 5532 bp in length  
\* 57673 57772: gap of unknown length  
\* 57773 62313: contig of 4541 bp in length  
\* 62314 62413: gap of unknown length  
\* 62414 66002: contig of 3589 bp in length  
\* 66003 66102: gap of unknown length  
\* 66103 72036: contig of 5934 bp in length  
\* 72037 72136: gap of unknown length  
\* 72137 77216: contig of 4980 bp in length  
\* 77217 80518: gap of unknown length  
\* 80518: contig of 3302 bp in length

80519 80618: gap of unknown length  
\* 80619 84810: contig of 4192 bp in length  
\* 84811 84910: gap of unknown length  
\* 84911 89660: contig of 4750 bp in length  
\* 89661 89760: gap of unknown length  
\* 89761 95933: contig of 6173 bp in length  
\* 95934 96033: gap of unknown length  
\* 96034 100988: contig of 4955 bp in length  
\* 100989 101088: gap of unknown length  
\* 101089 106799: contig of 5711 bp in length  
\* 106800 106899: gap of unknown length  
\* 106900 115241: contig of 8342 bp in length  
\* 115242 115341: gap of unknown length  
\* 115342 121457: contig of 6116 bp in length  
\* 121458 121557: gap of unknown length  
\* 121558 128525: contig of 6968 bp in length  
\* 128526 128625: gap of unknown length  
\* 128626 136102: contig of 7477 bp in length  
\* 136103 136202: gap of unknown length  
\* 136203 143456: contig of 7254 bp in length  
\* 143457 143557: gap of unknown length  
\* 143557 152354: contig of 8798 bp in length.  
Location/Qualifiers  
1. 152354

FEATURES  
source 1. /organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 45796 a 28744 c 29399 g 44369 t 4046 others  
ORIGIN

## Alignment Scores:

Pred. No.: 0.000623 Length: 152354  
Score: 162.50 Matches: 42  
Percent Similarity: 42.75% Conservative: 17  
Best Local Similarity: 30.43% Mismatches: 50  
Query Match: 22.82% Indels: 29  
DB: 2 Gaps: 5

US-09-997-610-2\_COPY\_18\_149 (1-132) x AC120598 (1-152354)

QY 2 GLYPRFADLHNSIPROARGPROGILUGLVALGILYPRPROGLYALAPROGLYLEU 21

DB 84076 GGCCAGCGTGGCTTCCAGAGACCCCGAGGCGCTCCAGAGACTCCAGAGACCCCGAGCTGTG 84135

QY 22 -----ProGlnTyrThrGlyGluLe----- 28

DB 84136 ATGCCCCCTACACACACACCCAGGAGATGTCACAGATATGGGCTGGATTCAT 84195

QY 29 -----SerGluMetThrIlyscYsProGysProAspIle 39

DB 84196 GGCCTGAACCCCGCATGCTACGGGGCTAAGAAAGCAAGATGGAGGGCAGCCTAT 84255

QY 40 GLUARGSERALAPHETHRVALYLSLEUSERGILYLSLEUPROLEUPROPHELYPROILE 59

DB 84256 GAGATGCGTGCATTATCCGCGGAGACTAACCGACCTTTCCACCGGTGGGGCCCGCATG 84315

QY 60 ILEPETHRGLYALAEUARYASNLAGLARGASPLEULYSGILAJAMEGLYVALPHE 79

DB 84316 AAGTTTAAACAACGCTGTATACGGCAGACAGAACCCGACAGACGATCTTC 84375

QY 80 ALACYSARGVALAPROGLYASNTYTYTSERSESRHEASRYVALISLUENHLSHCYLYLS 99

DB 84376 ACCTGTGAGGCGCTCGGTCTACTACTTTCACATACACAGTT-----CACTGCAG 84426

QY 100 ---ValasnlIeTPrLeuMeTargLysGlnIleLeuAlaAsnLysGluIle 116

DB 84427 GGGGCAACGtGTGGTT-----GCTCTATTCAAGAACACGAGCCCGTG 84471

Search completed: February 19, 2003, 23:01:51  
Job time: 1969.21 secs

















GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 16:09:30 : Search time 5.34727 Seconds  
(without alignments)  
4645.518 Million cell updates/sec

Title: US-09-997-610-1\_COPY\_56\_136

Perfect score: 81  
Sequence: 1 ggacctcagcagcaccaccag.....caccatatacagagaata 81

Scoring table: IDENTIFY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/lna/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/lna/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/lna/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/lna/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/lna/PCFUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/lna/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30.4	37.5	3394	1	US-08-159-784-4
2	29.8	36.8	1881	4	US-09-029-348-20
3	29.4	36.3	821	4	US-09-342-681C-7
4	29.4	36.3	1176	4	US-09-342-681C-14
5	29.4	36.3	1574	4	US-09-342-681C-1
6	26.8	33.1	390	4	US-09-134-001C-635
7	26.8	33.1	513	4	US-09-134-001C-647
8	26.8	33.1	585	4	US-09-134-001C-726
9	26.8	33.1	810	4	US-09-134-001C-624
10	26.8	33.1	3552	4	US-09-134-001C-693
11	26.6	32.8	1588	6	5510466-3
12	26.6	32.8	1807	6	5510466-1
13	26.6	32.8	1686	6	5510466-1
14	26.2	32.3	1794	5	PCT-US95-14418-3
15	26.2	32.3	1794	5	PCT-US95-15327-3
16	26.2	32.3	2167	4	US-09-067-351-5
17	26.2	32.3	2167	4	US-09-360-490-5
18	26.2	32.3	2504	1	US-08-073-384C-3
19	26.2	32.3	2504	1	US-08-254-359A-3
20	26.2	32.3	2504	1	US-08-483-043-3
21	26.2	32.3	2504	1	US-08-481-238-3
22	26.2	32.3	2504	2	US-08-471-066B-3
23	26.2	32.3	2504	2	US-08-484-956-3
24	26.2	32.3	2504	2	US-08-757-653-3
25	26.2	32.3	2504	2	US-08-599-491-3
26	26.2	32.3	2504	2	US-08-756-386-3
27	26.2	32.3	2504	2	US-08-823-516-3

28	26.2	32.3	2504	3	US-08-682-853A-3	Sequence 3, Appl1
29	26.2	32.3	2504	3	US-08-759-038-3	Sequence 3, Appl1
30	26.2	32.3	2504	3	US-08-758-314-3	Sequence 3, Appl1
31	26.2	32.3	2504	4	US-09-350-309-3	Sequence 3, Appl1
32	26.2	32.3	2504	4	US-08-520-946-3	Sequence 3, Appl1
33	26.2	32.3	2505	1	US-07-977-434-9	Sequence 9, Appl1
34	26.2	32.3	2505	1	US-08-458-819-9	Sequence 9, Appl1
35	26.2	32.3	2505	5	PCT-US91-07035-9	Sequence 9, Appl1
36	26.2	32.3	2640	1	US-08-384-490-30	Sequence 30, Appl1
37	26.2	32.3	2640	1	US-08-459-383-30	Sequence 30, Appl1
38	26.2	32.3	3048	5	PCT-US95-14418-1	Sequence 1, Appl1
39	26.2	32.3	3048	5	PCT-US95-15327-1	Sequence 1, Appl1
40	26.2	32.3	5228	4	US-09-428-711A-15	Sequence 15, Appl1
41	26.2	32.3	5102	1	US-08-494-168-1	Sequence 1, Appl1
42	25.8	31.9	900	4	US-09-198-603C-1	Sequence 1, Appl1
43	25.6	31.6	728	4	US-09-336-536-2	Sequence 2, Appl1
44	25.6	31.6	1074	2	US-08-627-151A-15	Sequence 15, Appl1
45	25.6	31.6	1203	4	US-09-086-010-1	Sequence 1, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-159-784-4  
; Sequence 4, Application US/08159784  
; Patent No. 5643783  
; GENERAL INFORMATION:  
; APPLICANT: Bjorn R. Olsen  
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; COMPUTER: IBM PS/2 Model 50z or 55sx  
; OPERATING SYSTEM: MS-DOS (Version 5.0)  
; SOFTWARE: Wordperfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/159,784  
; FILING DATE: December 1, 1993  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: John F. Freeman  
; REGISTRATION NUMBER: 29,066  
; REFERENCE/DOCKET NUMBER: 00246/170001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3394  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-159-784-4  
Query Match 37.5%; Score 30.4; DB 1; Length 3394;  
Best Local Similarity 71.4%; Pred. No. 0.42; 16; Indels 0; Gaps 0;  
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;  
QY 1 GGACCTCCAGCAGACCCAGCCAGAGAGTGGGCTCTGTCGACGAG 56  
DB 670 GGCGCTCAGAGACCCAGGGGCCCAAGAGAGTGGGCCCCCGGACGACGAG 725

```

RESULT 2
US-09-029-348-20
: Sequence 20, Application US/09029348
: Patent No. 6171827
: GENERAL INFORMATION:
: APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
: TITLE OF INVENTION: NOVEL PROCOLLAGENS
: FILE REFERENCE: d087857PUS LISTING
: CURRENT APPLICATION NUMBER: US/09/029.348
: CURRENT FILING DATE: 1998-05-07
: NUMBER OF SEQ ID NOS: 20
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 20
: LENGTH: 1881
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: SEQUENCED
: OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS
US-09-029-348-20

```

[illegible]

RESULT 3  
 US-09-342-681C-7  
 : Sequence 7, Application US/09342681C  
 : Patent No. 6355782  
 : GENERAL INFORMATION:  
 : APPLICANT : Zonana et al.  
 : TITLE OF INVENTION : Hypohydrotic ectodermal dysplasia genes and proteins  
 : FILE REFERENCE : 52978  
 : CURRENT APPLICATION NUMBER : US/09/342,681C  
 : CURRENT FILING DATE : 1999-06-29  
 : PRIOR APPLICATION NUMBER : 60/092,279  
 : PRIOR FILING DATE : 1998-07-09  
 : PRIOR APPLICATION NUMBER : 60/112,366  
 : PRIOR FILING DATE : 1998-12-15  
 : NUMBER OF SEQ ID NOS : 123  
 : SOFTWARE : PatentIn Ver. 2.1  
 : SEQ ID NO 7  
 : LENGTH: 821  
 : TYPE: DNA  
 : ORGANISM: Homo sapiens  
 US-09-342-681C-7

Query Match	36.3%	Score 29.4	DB 4	Length 821
Best Local Similarity	60.8%	Pred. NO. 0.62		
Matches	48	Conservative	0	Mismatches 31; Indels 0; Gaps 0
QY	1	GGACCTCCAGCAGCACCCAGCGCCCCAGAGAAGTGGGGCTCTGTGTCACACAGTTTA	60	
Db	319	GGCCCTCCAGAGACCCCCAGAGACTCTCCAGAGACCCCGAGCGACCCCGAAGATTCCAGGATT	378	
QY	61	CCACAATATACAGGAGAAA	79	
Db	379	CCTGGAATTCAGGAGACAA	397	

RESULT 4  
US-09-342-681C-14

Sequence 14, Application US/09342681C  
 Patent No. 6355782  
 GENERAL INFORMATION:  
 APPLICANT: Zonana et al.  
 TITLE OF INVENTION: Hypohydrotic ectodermal dysplasia genes and proteins  
 FILE REFERENCE: 52978  
 CURRENT APPLICATION NUMBER: US/09/342,681C  
 CURRENT FILING DATE: 1999-06-29  
 PRIOR APPLICATION NUMBER: 60/092,279  
 PRIOR FILING DATE: 1998-07-09  
 PRIOR APPLICATION NUMBER: 60/112,366  
 PRIOR FILING DATE: 1998-12-15  
 NUMBER OF SEQ ID NOS: 123  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 14  
 LENGTH: 1176  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1)..(1176)  
 US-09-342-681C-14

	Query Match	Score 36.3%	DB 29.4	Length 1176
	Best Local Similarity	Pred. 60.8%	No. 0.68	
Matches	48	Conservative	0	Mismatches 31; Indels 0; Gaps 0;
QY	1	GCACCTCAGACACACCCAGCGGCCACAGAAAGTGGGGCCCTCGTGCACCCAGCTTGA	60	
DB	556	GGCCCTCCAGACGCCCCAGACCTCCAGAACCCCGACGACCCCGACGAATTCAGAGATT	615	
QY	61	CCACAAATATACGAGGAAA	79	
DB	616	CCTGGAATTCGAGGAACAA	634	

```

RESULT 5
US-09-342-681C-1
Sequence 1, Application US/09342681C
Patent No. 6355782
GENERAL INFORMATION:
APPLICANT: Zonana et al.
TITLE OF INVENTION: Hypohydrotic ectodermal dysplasia genes and proteins
FILE REFERENCE: 52978
CURRENT APPLICATION NUMBER: US/09/342,681C
CURRENT FILING DATE: 1998-06-29
PRIOR APPLICATION NUMBER: 60/092,279
PRIOR FILING DATE: 1998-07-09
PRIOR APPLICATION NUMBER: 60/112,366
PRIOR FILING DATE: 1998-12-15
NUMBER OF SEQ ID NOS: 123
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 1574
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (242)..(1417)
US-09-342-681C-1

```

	Query Match	36.3%;	Score 29.4;	DB 4;	Length 1574;
	Best Local Similarity	60.8%;	Pred. No. 0.74;		
	Matches   48; Conservative		Mismatches   31;	Indels         0;	Gaps           0;
Oy	1 GGACCTTCAGCAGACCCCCGAGGCCCCCACAAGAATGGGGGCCTGTGGTGAACCAAGTTTA	60			
Dd	797 GGCCTTCAGGAGCCCCCGAGACTTCACAGACCCCGAGAACCCCGAGAAATTCCAAGGANT	856			
Oy	61 CCACAATAATACAGGAAAAA	79			
Dd	857 CCTGGAATTCAGGAGACAA	875			

Db 857 CCTGGAATTCACGAACAA 875

```
RESULT 6
US-09-134-001C-635/C
; Sequence 635, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 635
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-635

Query Match
Best Local Similarity 33.1%; Score 26.8; DB 4; Length 390;
Matches 46; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

OY 1 GGACCTCCAGCACACCCCGCCGAGAAAGTGGGCTCTCTGTGCACCAAGTTTA 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 159 GGTACCCCGCAGAACAGTACGCCAGCAACCAAGTAACCAACGACGAGTACG 100
OY 61 CCACAATATACAGAGAA 78
DB 99 CCAGCAGAACCAAGTAAA 82

RESULT 7
US-09-134-001C-647/C
; Sequence 647, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 647
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-647

Query Match
Best Local Similarity 33.1%; Score 26.8; DB 4; Length 513;
Matches 46; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

OY 1 GGACCTCCAGCACACCCCGCCGAGAAAGTGGGCTCTCTGTGCACCAAGTTTA 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 298 GGTACCCCGCAGAACAGTACGCCAGCAACCAAGTAACCAACGACGAGTACG 239
OY 61 CCACAATATACAGAGAA 78
DB 238 CCAGCAGAACCAAGTAAA 221

RESULT 8
US-09-134-001C-726
```

```
; Sequence 726, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 726
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-726

Query Match
Best Local Similarity 33.1%; Score 26.8; DB 4; Length 585;
Matches 46; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

OY 1 GGACCTCCAGCACACCCCGCCGAGAAAGTGGGCTCTCTGTGCACCAAGTTTA 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 270 GGTACCCCGCAGAACAGTACGCCAGCAACCAAGTAACCAACGACGAGTACG 329
OY 61 CCACAATATACAGAGAA 78
DB 330 CCAGCAGAACCAAGTAAA 347

RESULT 9
US-09-134-001C-624/C
; Sequence 624, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 624
; LENGTH: 810
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-624

Query Match
Best Local Similarity 33.1%; Score 26.8; DB 4; Length 810;
Matches 46; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

OY 1 GGACCTCCAGCACACCCCGCCGAGAAAGTGGGCTCTCTGTGCACCAAGTTTA 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 266 GGTACCCCGCAGAACAGTACGCCAGCAACCAAGTAACCAACGACGAGTACG 207
OY 61 CCACAATATACAGAGAA 78
DB 206 CCAGCAGAACCAAGTAAA 189

RESULT 10
US-09-134-001C-693
; Sequence 693, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
```

; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 693  
; LENGTH: 3552  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-693

Query Match 33.1%; Score 26.8; DB 4; Length 3552;  
Best Local Similarity 59.0%; Pred. No. 6.5;  
Matches 46; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 GGACCTCCAGCACACCCGCCCGAGAGAACTGGGCTCTGCTGCACAGTTTA 60  
Db 3166 GGTACGCCAGAGAACAGGTACGCCAGACAGTAACCAACGGAACGAGTACG 3225

QY 61 CCACATATACAGAGAA 78  
Db 3226 CCAGCAGAACCGAGTAA 3243

RESULT 11  
5510466-3  
; Patent No. 5510466  
; APPLICANT: KREIGER, MONTY; KODAMA, TATSUHIKO  
; TITLE OF INVENTION: SCAVENGER RECEPTOR PROTEIN AND ANTIBODY  
; THERETO  
; NUMBER OF SEQUENCES: 12  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/307,400  
; FILING DATE: 16-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 997,113  
; FILING DATE: 24-DEC-1992  
; APPLICATION NUMBER: 391,486  
; FILING DATE: 09-AUG-1989  
; APPLICATION NUMBER: 272,002  
; FILING DATE: 15-NOV-1988  
; SEQ ID NO: 3  
; LENGTH: 1588  
5510466-3

Query Match 32.8%; Score 26.6; DB 6; Length 1588;  
Best Local Similarity 58.0%; Pred. No. 6.1;  
Matches 47; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1 GGACCTCCAGCACACCCGCCCGAGAGAACTGGGCTCTGCTGCACAGTTTA 60  
Db 817 GGTCTCTCTGACCTCGAGTGAAGAGATGAGGCCCTCTGCAAAATGTATA 876

QY 61 CCACATATACAGAGAA 81  
Db 877 CCAGCTTTCAGGTCTAATA 897

RESULT 12  
5510466-1  
; Patent No. 5510466  
; APPLICANT: KREIGER, MONTY; KODAMA, TATSUHIKO  
; TITLE OF INVENTION: SCAVENGER RECEPTOR PROTEIN AND ANTIBODY  
; THERETO  
; NUMBER OF SEQUENCES: 12  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/307,400

; FILING DATE: 16-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 997,113  
; FILING DATE: 24-DEC-1992  
; APPLICATION NUMBER: 391,486  
; FILING DATE: 09-AUG-1989  
; APPLICATION NUMBER: 272,002  
; FILING DATE: 15-NOV-1988  
; SEQ ID NO: 1  
; LENGTH: 1807  
5510466-1

Query Match 32.8%; Score 26.6; DB 6; Length 1807;  
Best Local Similarity 58.0%; Pred. No. 6.3;  
Matches 47; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1 GGACCTCCAGCACACCCGCCCGAGAGAACTGGGCTCTGCTGCACAGTTTA 60  
Db 849 GGTCTCTGACCTCGAGTGAAGAGATGAGGCCCTCTGCAAAATGTATA 908

QY 61 CCACATATACAGAGAA 81  
Db 909 CCAGCTTTCAGGTCTAATA 929

RESULT 13  
US-08-648-657-14  
; Sequence 14, Application US/08648657  
; Patent No. 5885813  
; GENERAL INFORMATION:  
; APPLICANT: Davis, Maria  
; APPLICANT: Moffett, R. Bruce  
; TITLE OF INVENTION: THERMOSTABLE DNA  
; TITLE OF INVENTION: POLYMERASES  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: Word Perfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/648,657  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; PRIOR APPLICATION DATA: including application one  
; PRIOR APPLICATION DATA: described below:  
; APPLICATION NUMBER: US 08/455,686  
; FILING DATE: May 31, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 219/304  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1686 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

FEATURE:  
NAME/KEY: FY4  
LOCATION: 1...1683  
US-08-648-657-14

Query Match 32.3%; Score 26.2; DB 2; Length 1686;  
Best Local Similarity 63.5%; Pred. No. 8.4;  
Matches 40; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 9 AGCACACCCAGCCCCAGAGAGTGGGCTCTGTGTGACACCAAGTTTACCAACATA 68  
DB 1487 AGCTCTTCCCGCCGCGGAGATGGGGCCCGCATGCTCTCCAGTCACGACGAGC 1546

QY 69 TAC 71  
DB 1547 TCC 1549

RESULT 14  
PCT-US95-14418-3

Sequence 3, Application PC/TUS9514418  
GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: DNA Encoding a Thermostable DNA Polymerase Enzyme

NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois

COUNTRY: United States of America

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/14418

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Gass, David A.

REGISTRATION NUMBER: 38,153

REFERENCE/DOCKET NUMBER: 28003/32330

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300

TELEFAX: 312/474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1794 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1794

PCT-US95-14418-3

Query Match 32.3%; Score 26.2; DB 5; Length 1794;  
Best Local Similarity 63.5%; Pred. No. 8.5;

Matches 40; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 9 AGCACACCCAGCCCCAGAGAGTGGGCTCTGTGTGACACCAAGTTTACCAACATA 68  
DB 1595 AGCTCTTCCCGCCGCGGAGATGGGGCCCGCATGCTCTCCAGTCACGACGAGC 1654

QY 69 TAC 71  
DB 1655 TCC 1657

RESULT 15

PCT-US95-15327-3

Sequence 3, Application PC/TUS9515327

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: Biologically Active Fragments of  
Thermus Flavus DNA Polymerase

NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois

COUNTRY: United States of America

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/15327

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Gass, David A.

REGISTRATION NUMBER: 38,153

REFERENCE/DOCKET NUMBER: 28003/31716

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300

TELEFAX: 312/474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1794 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1794

PCT-US95-15327-3

Query Match 32.3%; Score 26.2; DB 5; Length 1794;  
Best Local Similarity 63.5%; Pred. No. 8.5;  
Matches 40; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 9 AGCACACCCAGCCCCAGAGAGTGGGCTCTGTGTGACACCAAGTTTACCAACATA 68  
DB 1595 AGCTCTTCCCGCCGCGGAGATGGGGCCCGCATGCTCTCCAGTCACGACGAGC 1654

QY 69 TAC 71  
DB 1655 TCC 1657

Search completed: February 19, 2003, 22:58:45  
Job time : 9.34727 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 15:08:07 ; Search time 28.0336 Seconds  
(without alignments)  
6506.903 Million cell updates/sec

Title: US-09-997-610-1\_COPY\_56\_136

Perfect score: 81

Sequence: 1 ggaacctccagcacaccacccag.....cacaatatacagagaata 81

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	48.1	1338	24	Human genset metab
2	32	39.5	6512	24	Mouse ischaemic co
3	32	39.5	61710	22	Human immune/haema
4	31.2	38.5	5010	24	Human Tumour Endot
5	31.2	38.5	6436	22	Human Polynucleoti
6	31.2	38.5	51935	22	Human Immune/haema
7	30.4	37.5	549	22	CDNA #49 encoding
8	30.4	37.5	1317	22	Murine HSP47 inter
9	30.4	37.5	3394	18	Human alpha-1 col1

10	30.4	37.5	3394	20	AAx78379	Human alpha1 (XVII
11	30.4	37.5	3394	24	ABN95680	Gene #2178 used to
12	30.2	37.3	5467	22	AAH98343	Human EST-derived
13	30.2	37.3	5468	22	AAH98411	Human EST-derived
14	30	37.0	707	22	AA544890	Human contig Polyn
15	30	37.0	2686	22	AA527013	CDNA encoding nove
16	30	37.0	3321	22	AA544718	Human full-length
17	29.8	36.8	561	24	ABK44107	CDNA #47 encoding
18	29.8	36.8	617	24	ABK44111	CDNA #51 encoding
19	29.8	36.8	648	24	ABK44126	CDNA #56 encoding
20	29.8	36.8	653	24	ABK44113	CDNA #53 encoding
21	29.8	36.8	655	24	ABK44128	CDNA #68 encoding
22	29.8	36.8	656	24	ABK44129	CDNA #69 encoding
23	29.8	36.8	658	24	ABK44131	CDNA #71 encoding
24	29.8	36.8	659	24	ABK44112	CDNA #52 encoding
25	29.8	36.8	780	24	ABK44065	CDNA #5 encoding h
26	29.8	36.8	823	24	ABK44088	CDNA #28 encoding
27	29.8	36.8	994	22	AAE44902	Human breast cance
28	29.8	36.8	1881	18	AAE59892	Coding sequence fo
29	29.8	36.8	4428	22	AAE06574	Bovine alpha1(III)
30	29.8	36.8	4428	22	AAE06575	Porcine alpha1(III)
31	29.8	36.8	4428	22	AAE06578	Vector PAC3A1 cont
32	29.8	36.8	5460	17	AAE16508	Collagen type III
33	29.8	36.8	5460	22	ABK83117	Human Tumour Endot
34	29.8	36.8	5460	24	ABE92101	Human EDAl-II exon
35	29.8	36.8	5466	23	AA579378	DNA encoding novel
36	29.8	36.8	9287	24	ABK64501	Human benign prost
37	29.6	36.5	1619	22	AA527365	CDNA encoding nove
38	29.6	36.5	1619	22	AA534830	CDNA encoding nove
39	29.6	36.5	6674	22	AA158233	Human polynucleoti
40	29.4	36.3	821	24	ABE51013	Human EDAl-II exon
41	29.4	36.3	1176	24	ABE51020	Human EDAl-II open
42	29.4	36.3	1574	24	ABE51009	Human EDAl-II enco
43	29.4	36.3	4062	22	AAH33347	Human colon cancer
44	29.4	36.3	9192	22	ABE92117	Human Tumour Endot
45	29.4	36.3	9483	22	AA544690	Human full-length

#### ALIGNMENTS

RESULT 1	AAE44066	AAE44066 standard; CDNA: 1338 BP.
ID	AAE44066	
AC	AAE44066;	
XX		
DT	27-SEP-2002 (first entry)	
XX		
DE	Human genset metabolic gene (GMC-9) CDNA sequence.	
XX		
KW	Human; gene; ss; gene therapy; genset metabolic gene; GMC-7A; GMC-7B;	
KW	GMC-8; GMC-9; GMC-10; GMC-11; metabolic-related disorder; obesity;	
KW	Impaired glucose tolerance; insulin resistance; Syndrome X;	
KW	Type II diabetes; hyperlipidaemia; atherosclerosis; hypertension;	
KW	heart disease; cardiac insufficiency; coronary insufficiency;	
KW	high blood pressure; insulin sensitizer;	
XX	non-insulin dependent diabetes mellitus.	
XX		
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	1..1338
FT		/*tag= a
FT		/partial
FT		/product= "Human GMC-9 protein"
FT		/note= "No stop codon is given"
PN	WO200255694-A2.	
XX		
PD	18-JUL-2002.	
XX		
PF	15-JAN-2002; 2002WO-1B01215.	

```
XX 16-JAN-2001; 2001US-26235P.
PR
XX (GEST ) GENSET.
XX
XX Erickson MR, Bour BA, Bihain B, Tanaka H;
XX
XX WPI; 2002-557821/59.
XX
XX P-PSDB; AAO15423.
XX
XX Treating or preventing a metabolic-related disease or disorder, e.g.
XX obesity, impaired glucose tolerance, insulin resistance, Syndrome X, or
XX Type II diabetes, comprises administering Genset Metabolic Genes
XX
XX Disclosure; Page 122-124; 128pp; English.
XX
XX The invention comprises the amino acid and coding sequences of six human
XX genset metabolic genes (GMG-7A, GMG-7B, GMG-8, GMG-9, GMG-10 and GMG-11).
XX
XX The GMG DNA and protein sequences of the invention are useful for
XX treating or preventing metabolic-related disorders, such as: obesity;
XX impaired glucose tolerance; insulin resistance; Syndrome X; Type II
XX diabetes; hyperlipidaemia; atherosclerosis; hypertension; and heart
XX diseases (e.g. cardiac insufficiency, coronary insufficiency or high
XX blood pressure). The GMG DNA and protein sequences of the invention may
XX also be used as insulin sensitizers - for improving insulin sensitivity
XX in persons with non-insulin dependent diabetes mellitus. The present cDNA
XX sequence encodes the human GMG-9 protein.
XX
XX Sequence 1338 BP; 359 A; 288 C; 350 G; 341 T; 0 other;
SQ
Query Match 48.1%; Score 39; DB 24; Length 1338;
Best Local Similarity 89.4%; Pred. No. 0.0035;
Matches 42; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
XX
XX 35 TGGGGCCTCTGTCACAGGTTTACCAATATACAGGAGAAATA 81
XX ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 50 TCGAATCTCTGTGGTCCCGAGTTTACCACATATACAGGAGAAATA 96
XX
XX RESULT 2
XX AB199819
XX ID AB199819 standard; cDNA; 6512 BP.
XX
XX AC AB199819;
XX
XX 07-MAR-2002 (first entry)
XX
XX Mouse ischaemic condition related cDNA sequence SEQ ID NO:932.
XX
XX DE Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
XX KW vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
XX
XX OS Mus musculus.
XX
XX PN WO200188188-A2.
XX
XX 22-NOV-2001.
XX
XX PD 18-MAY-2001; 2001WO-JP04192.
XX
XX PF 18-MAY-2001; 2001WO-JP04192.
XX
XX PR 18-MAY-2000; 2000JP-0145977.
XX
XX (UYNF-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX
XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
XX
XX WPI; 2002-034733/04.
XX
XX P-PSDB; ABB57334.
XX
XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring
XX PT expression levels of particular genes defined in the specification or
XX PT by determining the expression profile of a gene group comprising these
XX PT genes -
```

```
XX Claim 2; Page 2340-2352; 2690pp; English.
XX
XX The present invention describes a method for examining ischaemic
XX conditions, comprising measuring the expression levels of particular
XX CC genes (1) in a test sample or determining the expression profile of a
XX CC gene group in the sample comprising genes selected from (1). The method
XX CC is useful for examining the ischaemic condition (e.g. compressive
XX CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
XX CC expression levels of particular genes (AB199202 to AB199912, encoding
XX CC the protein sequences in ABB57020 to ABB57374) or by determining the
XX CC expression profile of a gene group comprising these genes. The
XX CC expression levels or expression profiles produced by these genes are
XX CC used as an indicator when screening for ischaemic condition-improving
XX CC drugs or therapeutics for ischaemic diseases. AB199913 and AB199914
XX CC represent PCR primers for a mouse ischaemic condition related sequence,
XX CC which are used in the exemplification of the present invention.
XX
XX Sequence 6512 BP; 1553 A; 1798 C; 1950 G; 1211 T; 0 other;
SQ
Query Match 39.5%; Score 32; DB 24; Length 6512;
Best Local Similarity 62.5%; Pred. No. 1;
Matches 50; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
XX
XX 1 GGACCTCCAGCACACCCAGGCCCCAGAGAAAGTGGGCTCTGTGCACACGATTTA 60
XX ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 2694 GGACAGTCAGGCCCTCTGCTCTGTGCACAGCAGGAGACCTGGAGTTCCAGGGTTTC 2753
XX
XX 61 CCACATATACAGGAGAAAT 80
XX
XX 2754 CCAGGTTCTAAAGGTGAAT 2773
XX
XX RESULT 3
XX AAK83782/c
XX ID AAK83782 standard; DNA; 61710 BP.
XX
XX AC AAK83782;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38594.
XX
XX DE Human immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO200157182-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX PR 04-FEB-2000; 2000US-0180628.
XX
XX PR 24-FEB-2000; 2000US-0184664.
XX
XX 02-MAR-2000; 2000US-0186350.
XX
XX 16-MAR-2000; 2000US-0189874.
XX
XX 17-MAR-2000; 2000US-0190076.
XX
XX 18-APR-2000; 2000US-0198123.
XX
XX 19-MAY-2000; 2000US-0205515.
XX
XX 07-JUN-2000; 2000US-0209467.
XX
XX 28-JUN-2000; 2000US-0214886.
XX
XX 30-JUN-2000; 2000US-0215135.
XX
XX 07-JUL-2000; 2000US-0216647.
XX
XX 07-JUL-2000; 2000US-0216880.
XX
XX PR 11-JUL-2000; 2000US-0217487.
XX
XX 11-JUL-2000; 2000US-0217496.
XX
XX PR 14-JUL-2000; 2000US-0218290.
XX
XX 26-JUL-2000; 2000US-0220963.
XX
XX PR 26-JUL-2000; 2000US-0220964.
XX
XX 14-AUG-2000; 2000US-0224518.
```



CC	represent sequences used in the exemplification of the present invention.
XX	
SQ	Sequence 61710 BP; 11560 A; 18542 C; 18148 G; 13460 T; 0 other;
	Query Match                 39.5%; Score 32; DB 22; Length 61710;
	Best Local Similarity   65.3%; Pred No. 18;
	Matches      47; Conservative   0; Mismatches   25; Indels    0; Gaps      0
OY	1 GGACCTCAGCACACCAGCCCCAGAAGATGGGCGTCCTGTGCACCAAGTTTA 60   44831 GGCCCTCAGAGACCAGCGGCCAAGAAGATGGGCGCCCGGACACACAGTAG 44772
OY	61 CCACAATTACAA 72   Db 44771 CAACTCTGACAA 44760
RESULT 4	
ID	ABL92114 standard; cDNA; 5010 BP.
XX	ABL92114;
XX	
DE	30-MAY-2002 (first entry)
XX	
DE	Human Tumour Endothelial Marker polynucleotide SEQ ID NO 251.
KW	Human; mouse; rat; TEM; tumour endothelial marker; NEM; PEM; cytostatic; normal endothelial marker; pan-endothelial marker; immunostimulant; antiangiogenic; tumour; neovascularization; vascularised tumour; polycystic kidney disease; diabetes; retinopathy; rheumatoid arthritis; psoriasis; gene; ss.
OS	Homo sapiens.
XX	
PN	WO200210217-A2.
PD	07-FEB-2002.
PF	01-AUG-2001; 2001MO-US24031.
PR	02-AUG-2000; 2000US-22259P. 11-AUG-2000; 2000US-22436OP. 11-APR-2001; 2001US-282850P.
PA	(UYJO ) UNIV JOHNS HOPKINS.
PL	St Croix B, Kinzler KW, Vogelstein B;
DR	WPJ: 2002-291856/33.
P	P-PSDB; ABB90760.
PT	An isolated molecule comprising an antibody variable region which specifically binds to an extracellular domain of a tumor endothelial marker (TEM) protein, useful for inhibiting tumor growth -
PS	Claim 58: Page 245-246; 331pp; English.
XX	The invention relates to an isolated molecule comprising an antibody variable region which specifically binds to an extracellular domain of a tumor endothelial marker (TEM) protein selected from ABB90732, ABB90740, ABB90749, ABB90750 and ABB90769. The antibodies which bind to TEM proteins have cytostatic, immunostimulant and antiangiogenic activity. They are useful for inhibiting tumor growth, neovascularisation in subjects bearing a vascularized tumour, polycystic kidney disease, diabetic retinopathy, rheumatoid arthritis and psoriasis. Human, mouse and rat TEM genes and the encoded proteins (ABL92114 and ABB90721-ABB90789) are disclosed, as are marker oligonucleotides and sequences: tumour endothelial markers (TEM) ABL91996-ABL92041 and ABL92143-ABL92191; normal endothelial markers (NEM) ABL92042-ABL92074; and pan-endothelial markers (PEM) ABL91903-ABL91995.
XX	
XX	Sequence 5010 BP; 1157 A; 1354 C; 1658 G; 841 T; 0 other;

Query Match	38.5%	Score 31.2	DB 24	Length 5010
Best Local Similarity	66.2%	Pred. No. 1.8		
Matches 45	Conservative 0	Mismatches 23	Indels 0	Gaps 0
Qy 13	CACCCAGGGCCCCAGAGAAGTGGGGCTCTCTGTGCACACGATTTACACAAATATACA	72		
Db 1303	CAGCCCGACCTCCAGGTGACACGAGGTCTCTTGAAATTCAGGGCAGCCAGGATTTATA	1362		
Qy 73	GGAGAAT 80			
Db 1363	GGCGAAT 1370			
RESULT 5				
AA160019				
AA160019	standard; cDNA; 6436 BP.			
AC AA160019;				
DT 22-OCT-2001	(first entry)			
DE Human polynucleotide SEQ ID NO 4008.				
XX Human; noctropic; immunosuppressant; cytosatic; gene therapy; cancer;				
KM peripheral nervous system; neuropathy; central nervous system; CNS;				
KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;				
KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;				
KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;				
KM leukaemia; ss.				
XX Homo sapiens.				
XX Wo20015312-A1.				
XX 26-JUL-2001.				
XX 26-DEC-2000; 2000MO-US34263.				
XX 21-JAN-2000; 2000US-0488725.				
XX 25-APR-2000; 2000US-0552317.				
PR 09-JUL-2000; 2000US-0598042.				
PR 19-JUL-2000; 2000US-0620312.				
PR 03-AUG-2000; 2000US-0653450.				
PR 14-SEP-2000; 2000US-0662191.				
PR 19-OCT-2000; 2000US-0693036.				
PR 29-NOV-2000; 2000US-0727344.				
XX (HYSE-) HYSEQ INC.				
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;				
PI Wang Z, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;				
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;				
XX WPI: 2001-442253/47.				
DR P-PSDB; AAMA0863.				
PT Novel nucleic acids and polypeptides, useful for treating disorders				
PT such as central nervous system injuries -				
PS Claim 1; SEQ ID NO 4008; 10078pp; English.				
XX The invention relates to human nucleic acids (AA157798-AA161369) and				
CC the encoded polypeptides (AAW38642-AAW42213) with noctropic,				
CC immunosuppressant and cytosatic activity. The polynucleotides are useful				
CC in gene therapy. A composition containing a polypeptide or polynucleotide				
CC of the invention may be used to treat diseases of the peripheral nervous				
CC system, such as peripheral nervous injuries, peripheral neuropathy and				
CC localised neuropathies and central nervous system diseases, such as				
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic				
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the				
CC utilisation of the activities such as: Immune system suppression, haemo-				
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic				

CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.

XX Sequence 6436 BP; 1607 A; 1624 C; 1893 G; 1312 T; 0 other;

Query Match 38.5%; Score 31.2; DB 22; Length 6436;  
Best Local Similarity 66.2%; Pred. No. 1.9;  
Matches 45; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

OY 13 CACCCAGGCCCGCCAGAGAGTGGGCTCTGTGTCACAGTTTACCAATATACA 72

Db 1304 CAGCCCGGACCTCCAGTCACAGGCTCTCTCTGCAATTCACAGGACGAGATTATTA 1363

OY 73 GGAGAAAT 80

Db 1364 GCGGAAAT 1371

RESULT 6

AAK75883/c  
ID AAK75883 standard; DNA: 51935 BP.

XX AAK75883;

DT 07-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:30695.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

KW cytosolic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

XX MO200157182-A2.

PD 09-AUG-2001.

PF 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 11-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 14-JUL-2000; 2000US-0217496.

PR 26-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 14-AUG-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259578.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-483426/52.  
DR  
XX  
XX  
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -  
XX  
XX  
PS Disclosure; SEQ ID NO 30695; 3071pp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
XX Sequence 51935 BP; 14418 A; 12317 C; 11675 G; 13525 T; 0 other;  
Query Match 38.5%; Score 31.2; DB 22; Length 51935;  
Best Local Similarity 66.2%; Pred. No. 3.2;  
Matches 45; Conservative 0; Mismatches 23; Indels 0; Gaps 0;  
QY 13 CACCAGGAGCCAGAGAGTGGGGCTCTGTGTCACCAAGTTTACCAATATACA 72  
DB 46144 CACCCGAGACTCTCAGGTGACAGGAGTCTCTGGAATTCCAGGCGACCAAGATTATA 46085

QY 73 GGAGAAAT 80  
DB 46084 GGGCGAAAT 46077  
RESULT 7  
ID ABRK44109 standard; cDNA; 549 BP.  
XX  
XX ABRK44109;  
AC  
XX  
DT 21-MAY-2002 (first entry)  
XX  
XX cDNA #49 encoding human pancreatic tumour protein.  
XX  
XX Human; pancreatic tumour protein; immune response; pancreatic cancer;  
KM development of cancer; cancer progression; cytostatic; gene; ss.  
XX  
XX Homo sapiens.  
OS  
PN WO200212331-A2.  
PD 14-FEB-2002.  
XX  
XX 06-AUG-2001; 2001WO-US24619.  
PF  
XX  
XX 07-AUG-2000; 2000US-223130P.  
PR 30-JAN-2001; 2001US-265447P.  
PR 15-MAY-2001; 2001US-291201P.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Pyle RA, Xu J, Kalos MD;  
PI WPI; 2002-241741/29.  
DR  
XX  
XX Novel polynucleotide encoding pancreatic tumour polypeptides, useful in  
PT pharmaceutical compositions, e.g. vaccines, for treating pancreatic  
PT cancers -  
XX  
XX  
PS Claim 1; Page 126; 167pp; English.  
XX  
XX The present invention relates to the isolation of cDNA sequences  
CC encoding human pancreatic tumour proteins. The polynucleotide  
CC sequences encoding human pancreatic tumour proteins are useful for  
CC stimulating an immune response in a patient and treating pancreatic  
CC cancer in a patient. A host cell that expresses these polynucleotides  
CC is useful for determining the presence of cancer in a patient. A  
CC composition comprising the polynucleotide, its encoded protein, or an  
CC antibody that binds to the protein may be used in the diagnosis,  
CC prevention and/or treatment of diseases, particularly pancreatic  
CC cancer. The sequences of the invention are also useful in pharmaceutical  
CC compositions, e.g. vaccines, for the diagnosis and treatment of  
CC pancreatic cancer. Such compositions may be useful for inhibiting the  
CC development of cancer in a patient, or as markers for the progression  
CC of cancer. The polynucleotide sequences may also be used as probes  
CC or primers for nucleic acid hybridisation assays. ABRK44061-ABRK44209  
XX represent cDNA sequences encoding for human pancreatic tumour proteins.  
XX  
XX Sequence 549 BP; 122 A; 167 C; 137 G; 119 T; 4 other;  
Query Match 37.5%; Score 30.4; DB 24; Length 549;  
Best Local Similarity 60.5%; Pred. No. 1.8;  
Matches 49; Conservative 0; Mismatches 32; Indels 0; Gaps 0;  
QY 1 GGACTTCAGACACACCCAGCCGCCAGAGAGTGGGGCTCTGTGACACAGTTTA 60  
DB 406 GGATCTCCAGATACCAAGACCCCTGTGAACTGGGCAAGCTGTCTTACGGCCCT 465  
QY 61 CCACATATATACAGAGAAATA 81  
DB 466 CCAGGACTCTCTGNGCTATA 486

```

RESULT 8
AAH48067
ID AAH48067 standard; cDNA; 1317 BP.
XX
XX AAH48067;
AC
XX
XX 19-SEP-2001 (first entry)
DT
XX
DE Murine HSP47 interacting protein, #1, coding sequence.
XX
XX Murine; heat shock protein interacting protein; HSP47; ss.
XX
XX Mus sp.
OS
XX JP2001145493-A.
PN
XX
XX 29-MAY-2001.
PD
XX
XX 19-NOV-1999; 99JP-0330631.
PF
XX
XX 19-NOV-1999; 99JP-0330631.
PR
XX
XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
PA
XX
XX WPI: 2001-395263/42.
DR
XX P-PSDB: AAC64211.
DT
XX
XX Using the two-hybrid screening method to prepare proteins which
PT interact with the heat shock protein HSP47
PS
XX Claim 21; Page 18-19; 26pp; Japanese.
XX
XX The present invention relates to a method for preparing a protein which
CC interacts with the heat shock protein HSP47. The method involves the
CC two-hybrid screening method using the HSP47 gene and a mammalian cDNA
CC library. The present sequence is the coding sequence for a murine HSP47
CC interacting protein which was used in the present method. The HSP47
CC interacting proteins are useful for the diagnosis and treatment of
CC diseases caused by an increase or decrease in activity of HSP47.
XX
XX Sequence 1317 BP; 275 A; 422 C; 410 G; 210 T; 0 other:
SQ
Query Match 37.5%; Score 30.4; DB 22; Length 1317;
Best Local Similarity 71.4%; Pred. No. 2.3;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
OY 1 GGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGCTCTGTGCACGAG 56
DB 706 GGACGACCGGGGCGACGCGGCTCCTGGAGAGATGGGCGCCCTGTCCACGAG 761

```

```

XX
XX 01-DEC-1993; 93US-0159784.
PF
XX
XX 01-DEC-1993; 93US-0159784.
PR
XX
XX (HARD ) HARVARD COLLEGE.
PA
XX
XX Oh SP, Olsen BR;
PI
XX
XX WPI: 1997-350247/32.
DR
XX P-PSDB: AAM26327.
DT
XX
XX Nucleic acid encoding human alpha-1 collagen - for production of
PT recombinant alpha-1 collagen, for use in the treatment of cartilage
PT degeneration
XX
XX Claim 1; Column 23-30; 35pp; English.
PS
XX
XX A cDNA clone (AAH4484) codes for a human novel type alpha-1 (XVIII)
CC collagen (AAM26327) that is expressed in multiple tissues, especially
CC liver, lung and kidney. It was isolated from a placental cDNA
CC library using a probe based on an unidentified collagenous protein
CC and a probe based on mouse alpha-1 cDNA clone mc19. A claimed
CC plasmid comprising alpha-1 collagen nucleic acid and an expression
CC control sequence can be used to express recombinant collagen in
CC prokaryotic or eukaryotic (especially mammalian) host cells. The
CC collagen may be used to treat a patient suffering from a disease
CC associated with degradation of cartilage, and for supplementing
CC collagen.
XX
XX Sequence 3394 BP; 668 A; 1130 C; 1080 G; 516 T; 0 other:
SQ
Query Match 37.5%; Score 30.4; DB 18; Length 3394;
Best Local Similarity 71.4%; Pred. No. 2.9;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
OY 1 GGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGCTCTGTGCACGAG 56
DB 670 GGCCCTCCAGCACACCCAGGCCCCAGAGAAGTGGGCTCTGTGCACGAG 725

```

```

RESULT 10
AAH78379
ID AAH78379 standard; cDNA; 3394 BP.
XX
XX AAH78379;
AC
XX
XX 25-AUG-1999 (first entry)
DT
XX
DE Human alpha1 (XVIII) collagen cDNA.
XX
XX Alpha1(XVIII) collagen; mimetic; endostatin; atomic coordinate; library;
XX anti-angiogenic; heparin binding domain; receptor binding domain; mimetic;
XX alpha-helix A domain; carbohydrate recognition domain; CRD domain;
XX treatment; angiogenesis; tumour; human; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX CDS 1..2055
XX FT /*tag= a
XX FT /product= "alpha1(XVIII) collagen"
XX FT /note= "Partial sequence, no start codon given"
XX
XX WO931616-A1.
XX
XX 24-JUN-1999.
PD
XX
XX 16-DEC-1998; 98WO-US26783.
PF
XX
XX 16-DEC-1997; 97US-0069727.
PR
XX
XX (HARD ) HARVARD COLLEGE.
PA

```

```
XX
PI Honenester E, Olsen BR, Sasaki T, Timpl R;
XX
DR WPI; 1999-395243/33.
DR P-PSDB; AAY25113.
XX
PT Identifying mimetics of mammalian endostatin
XX
PS Disclosure; Fig 5A-C; 75pp; English.
XX
CC This invention describes a novel method for identifying mimetics of
CC mammalian endostatin. The method comprises identifying a compound
CC having atomic coordinates with non-trivial similarity to selected
CC coordinates of atoms of a mammalian endostatin involves (a) providing
CC a library of atomic coordinates of compounds in a library of candidate
CC compounds, (b) comparing the library of atomic coordinates to the
CC selected coordinates of a mammalian endostatin and (c) selecting from the
CC library at least one candidate compound on the basis of selection
CC criteria which include similarities between the atomic coordinates of the
CC selected candidate compound and the atomic coordinates of the mammalian
CC endostatin. The invention also describes the use of an anti-angiogenic
CC fragment of endostatin comprising a domain selected from a heparin
CC binding domain, a receptor binding domain, and exposed on alpha-helix A
CC domain, and a carbohydrate recognition domain (CRD) domain. The methods
CC can be used for designing and selecting endostatin mimics. The compounds
CC identified can be used for treating undesired angiogenesis, e.g. tumours.
CC This sequence encodes human alpha1(XVII) collagen which is used in the
CC description of the method.
XX
SQ Sequence 3394 BP; 668 A; 1130 C; 1080 G; 516 T; 0 other;

Query Match      37.5%; Score 30.4; DB 20; Length 3394;
Best Local Similarity 71.4%; Pred. No. 2.9;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY      1 GGACCTCCAGCACACCCAGGCCCCAGAGAGAGTGGGGCTCTGCTGTGCACGAG 56
        || ||||| || | ||| || ||||| || || || |||||
DB      670 GGCCCTCCAGAGCCCAAGGGCCCAAGAGAGAGTGGGGCCGCCCGGACACGAGG 725

RESULT 11
ABN95680
ID ABN95680 standard; DNA: 3394 BP.
XX
AC ABN95680.
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #2178 used to diagnose liver cancer.
XX
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
XX Homo sapiens.
XX
XX WO200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US30589.
XX
PR 02-OCT-2000; 2000US-237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX
DR WPI; 2002-426119/45.
XX
PT Diagnosing and detecting the progression of liver cancer,
PT hepatocellular carcinoma or metastatic liver tumor in a patient,
PT involves detecting the level of expression of two or more genes in a
```

```
PT liver tissue sample -
XX
PS Claim 1; SEQ ID NO 2178; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3394 BP; 668 A; 1130 C; 1080 G; 516 T; 0 other;

Query Match      37.5%; Score 30.4; DB 24; Length 3394;
Best Local Similarity 71.4%; Pred. No. 2.9;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY      1 GGACCTCCAGCACACCCAGGCCCCAGAGAGAGTGGGGCTCTGCTGTGCACGAG 56
        || ||||| || | ||| || ||||| || || || |||||
DB      670 GGCCCTCCAGAGCCCAAGGGCCCAAGAGAGAGTGGGGCCGCCCGGACACGAGG 725

RESULT 12
AAH98343
ID AAH98343 standard; CDNA: 5467 BP.
XX
AC AAH98343.
XX
DT 12-OCT-2001 (first entry)
XX
DE Human EST-derived coding sequence SEQ ID NO: 200.
XX
DE Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; ss.
XX
XX Homo sapiens.
XX
XX WO200154477-A2.
XX
PN 02-AUG-2001.
XX
PD 25-JAN-2001; 2001WO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
XX
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR WPI; 2001-476164/51.
XX
DR P-PSDB; AAM23684.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
PS Claim 1; Page 314-315; 1275pp; English.
XX
PT The present invention provides the protein and coding sequences of novel
```

CC proteins from a variety of organisms, including human, dog, cat, horse,  
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
CC from the organism of interest. They can be used in diagnostics,  
CC forensics, gene mapping, identification of mutations, to assess  
CC biodiversity and for nutritional purposes. The present sequence is a cDNA  
CC of the invention.  
XX  
SQ Sequence 5467 BP; 1330 A; 1323 C; 1550 G; 1264 T; 0 other;  
Query Match 37.3%; Score 30.2; DB 22; Length 5467;  
Best Local Similarity 62.7%; Pred. No. 3.8;  
Matches 47; Conservative 0; Mismatches 28; Indels 0; Gaps 0;  
DY 1 GGACCTCCAGCACACCCAGGCCCCAGAGAACTGGGCTCTGTGTCACAGGTTTA 60  
DB 634 GGCCCCCAGAGCCCGCCCGCCCTGTGATCATCTGTCCTGTCCTCCGATCT 693  
DY 61 CCACAATATACAGGA 75  
DB 694 CCAGGATACCAAGGA 708  
RESULT 13  
AAH98411  
ID AAH98411 standard; cDNA; 5468 BP.  
AC AAH98411;  
XX  
XX 12-OCT-2001 (first entry)  
DE Human EST-derived coding sequence SEQ ID NO: 268.  
XX  
XX Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
KW diagnostics; forensic test; gene mapping; genetic disorder;  
KW biodiversity; gene therapy; nutrition; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200154477-A2.  
PN  
XX  
XX 02-AUG-2001.  
PD  
XX  
XX 25-JAN-2001; 2001WO-US02687.  
PE  
XX  
XX 25-JAN-2000; 2000US-0491404.  
PR 17-JUL-2000; 2000US-0617746.  
PR 03-AUG-2000; 2000US-0631451.  
PR 15-SEP-2000; 2000US-0663870.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
PI Cao Y, Drmanac RA, Zhang J, Werhman T;  
PI  
XX  
XX WPI; 2001-476164/51.  
DR P-PSDB; AAM23752.  
DR  
XX  
XX Isolated polypeptide for treatment of diseases, diagnostics, raising  
PT antibodies and research use -  
XX  
XX Claim 1; Page 381-382; 1275pp; English.  
XX  
XX The present invention provides the protein and coding sequences of novel  
CC proteins from a variety of organisms, including human, dog, cat, horse,  
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
CC from the organism of interest. They can be used in diagnostics,  
CC forensics, gene mapping, identification of mutations, to assess  
CC biodiversity and for nutritional purposes. The present sequence is a cDNA  
CC of the invention.  
XX

SQ Sequence 5468 BP; 1330 A; 1323 C; 1551 G; 1264 T; 0 other;  
Query Match 37.3%; Score 30.2; DB 22; Length 5468;  
Best Local Similarity 62.7%; Pred. No. 3.8;  
Matches 47; Conservative 0; Mismatches 28; Indels 0; Gaps 0;  
DY 1 GGACCTCCAGCACACCCAGGCCCCAGAGAACTGGGCTCTGTGTCACAGGTTTA 60  
DB 634 GGCCCCCAGAGCCCGCCCGCCCTGTGATCATCTGTCCTGTCCTCCGATCT 693  
DY 61 CCACAATATACAGGA 75  
DB 694 CCAGGATACCAAGGA 708  
RESULT 14  
AAS44890  
ID AAS44890 standard; DNA; 707 BP.  
AC AAS44890;  
XX  
XX 18-DEC-2001 (first entry)  
DT  
XX  
XX Human contig polynucleotide sequence #143.  
DE  
XX  
XX Mammal: human; rhesus monkey; baker's yeast; fission yeast; Norway rat;  
KW mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukemia;  
KW cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;  
KW nervous system disorder; inflammatory disorder; cell differentiation; ds;  
KW angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;  
KW genetic disorder; bone regeneration; tendon; ligament; tissue repair;  
KW cytostatic; antirheumatic; antiarthritic; vulnery; antiinflammatory;  
KW antibacterial; immunosuppressive; vasotropic; antiparkinsonian;  
KW neuroprotective; osteopathic; antidiabetic; antiallergic; antiallergic;  
KW immunostimulant; analgesic; gene therapy.  
XX  
XX Homo sapiens.  
OS  
XX  
XX Synthetic.  
OS  
XX  
XX WO200164834-A2.  
PN  
XX  
XX 07-SEP-2001.  
PD  
XX  
XX 26-FEB-2001; 2001WO-US04926.  
PE  
XX  
XX 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.  
PR 17-JUN-2000; 2000US-0597707.  
PR 14-JUL-2000; 2000US-0616887.  
PR 19-SEP-2000; 2000US-0664641.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;  
PI Drmanac R;  
PI  
XX  
XX WPI; 2001-589862/66.  
DR P-PSDB; AAU27990.  
DR  
XX  
XX Novel polypeptides and nucleic acids obtained from cDNA libraries  
PT prepared from various human tissues, for diagnosis, treatment of  
PT cancer, neurological, inflammatory disorders and for use in arrays for  
PT detection -  
XX  
XX Claim 1; SEQ ID NO 487; 153pp; English.  
XX  
XX Sequences AAS44576-AAS44919 represent full-length polynucleotides and  
CC contig polynucleotides encoding polypeptides of the invention. The DNA  
CC and protein sequences are useful for the treatment, diagnosis and  
CC prevention of various types of disorder in a mammalian subject such as a  
CC human, dog, monkey, mouse, hamster or rat. The disorders include cancers  
CC such as leukemia, lymphoma and neuroblastoma, autoimmune disorders such



PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR MPI: 2001-465460/50.  
 DR P-PSDB; AAU17096.  
 XX  
 PT Novel polypeptides useful for diagnosing, treating, preventing and/or  
 PT prognosing disorders related to the proteins, including cancers, immune  
 PT disorders and neuronal disorders -  
 XX  
 PS Claim 1: SEQ ID No 48; 880pp; English.  
 XX  
 CC The invention relates to novel isolated polypeptides (I), and  
 CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for  
 CC diagnosing, preventing and treating diseases including immune system  
 CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune  
 CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ  
 CC transplant rejections and graft versus host disease, infectious diseases  
 CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and  
 CC other blood-related disorders (sickle cell anaemia), myeloproliferative

CC disorders, primary haematopoietic disorders, hyperproliferative  
 CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal  
 CC abnormalities (Down syndrome), ischemic injury (e.g. stroke), renal  
 CC disorders (e.g. glomerulonephritis), cardiovascular disorders  
 CC disorders (e.g. arrhythmia), respiratory disorders, dermatological disorders,  
 CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.  
 CC Addison's disease), reproductive system disorders, gastrointestinal  
 CC disorder (inflammatory disorders), liver disorders (cirrhosis),  
 CC as stimulators of B-cell responsiveness to pathogens, activators of  
 CC T-cells, to induce higher affinity antibodies, and as a means to induce  
 CC tumour proliferation in pathologies e.g. acquired immune deficiency  
 CC syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction  
 CC pathway protein coding sequences and PCR primers of the invention.  
 XX

Query Match 37.0%; Score 30; DB 22; Length 2686;

Best Local Similarity 61.5%; Pred. No. 3.7; Matches 48; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 1 GGACCTCCAGACACCCCGCCAGAGAGAGTGGGCTCTCTGTGACACAGTTTA 60

DB 1305 GGAGCTCTCTGCACACCGAGAGACCTATGTGAAGCGCTGACCTCTGACACGATTTT 1364

QY 61 CCACATATATACGAGGAA 78

DB 1365 CTGCACACGAGCTGACGGA 1382

Search completed: February 19, 2003, 22:56:13  
 Job time : 65.0336 secs

**THIS PAGE BLANK (USPTO)**



## JOURNAL

## COMMENT

Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquay@sanger.ac.uk  
Requests: clonerequest@sanger.ac.uk  
On Dec 13, 1999 this sequence version replaced g1:3164067.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.  
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at  
<http://www.sanger.ac.uk/projects/C-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr22>  
RP1-30209 is from the library RPCT-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pcypac2>  
This sequence is the entire insert of clone RP1-30209. The true left end of clone CTA-28282 is at 69682 in this sequence. The true right end of clone CTA-41562 is at 5167 in this sequence.

## FEATURES

source  
1. 145880  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="22"  
/clone="RP1-30209"  
/clone\_lib="RPCT-1"  
188. .245  
/note="MER3 repeat: matches 144. .209 of consensus"  
246. .571  
/note="AluX repeat: matches 1. .312 of consensus"  
572. .759  
/note="MER3 repeat: matches 1. .144 of consensus"  
783. .933  
/note="MER5A repeat: matches 26. .187 of consensus"  
1033. .1336  
/note="AluSp repeat: matches 1. .299 of consensus"  
1450. .1583  
/note="MIR repeat: matches 24. .160 of consensus"  
1687. .1752  
/note="L2 repeat: matches 2593. .2661 of consensus"  
2350. .2660  
/note="AluSc repeat: matches 3. .309 of consensus"  
2884. .2981  
/note="AluSq repeat: matches 2. .300 of consensus"  
3323. .3343  
/note="MLTIE repeat: matches 116. .136 of consensus"  
3344. .3652  
/note="AluY repeat: matches 1. .309 of consensus"  
3653. .3928  
/note="MLTIE repeat: matches 136. .359 of consensus"  
3929. .4278  
/note="THE1B repeat: matches 3. .364 of consensus"  
4279. .4485  
/note="MLTIE repeat: matches 359. .568 of consensus"  
5073. .5176  
/note="52 copies 2 mer ct 78 conserved"  
5181. .5491  
/note="AluJb repeat: matches 1. .311 of consensus"  
6369. .6485  
/note="L2 repeat: matches 2579. .2705 of consensus"

repeat\_region 6647. .6685  
/note="MADE1 repeat: matches 1. .23 of consensus"  
repeat\_region 6686. .6987  
/note="AluX repeat: matches 1. .302 of consensus"  
6988. .7036  
/note="MADE1 repeat: matches 23. .77 of consensus"  
7482. .7754  
/note="AluJb repeat: matches 9. .290 of consensus"  
7775. .8060  
/note="AluJo repeat: matches 1. .295 of consensus"  
8414. .8551  
/note="L2 repeat: matches 2553. .2706 of consensus"  
8914. .9030  
/note="MIR repeat: matches 147. .262 of consensus"  
9110. .9280  
/note="MIR repeat: matches 91. .262 of consensus"  
9283. .9412  
/note="MIR repeat: matches 15. .144 of consensus"  
9521. .9679  
/note="FAM repeat: matches 3. .161 of consensus"  
9820. .10225  
/note="MSTR repeat: matches 2. .425 of consensus"  
complement(10179. .10678)  
/note="match: GSS: Em:R56592"  
complement(10204. .10728)  
/note="match: GSS: Em:AQ701486"  
complement(10249. .10706)  
/note="match: GSS: Em:AQ225495"  
10312. .10383  
/note="MIR repeat: matches 79. .150 of consensus"  
10718. .11310  
/note="match: GSS: Em:B14024"  
10784. .11201  
/note="match: GSS: Em:B43656"  
11838. .11946  
/note="MIR repeat: matches 20. .137 of consensus"  
12174. .12445  
/note="L2 repeat: matches 1988. .2275 of consensus"  
12444. .12642  
/note="MIR repeat: matches 63. .241 of consensus"  
13017. .13369  
/note="match: SRS: Em:G49301"  
13331. .13397  
/note="MIR repeat: matches 174. .244 of consensus"  
13398. .13698  
/note="AluSp repeat: matches 1. .302 of consensus"  
13699. .13810  
/note="MIR repeat: matches 76. .174 of consensus"  
13806. .13919  
/note="MIR repeat: matches 77. .189 of consensus"  
13945. .14060  
/note="MIR repeat: matches 24. .142 of consensus"  
14061. .14367  
/note="AluY repeat: matches 1. .301 of consensus"  
14368. .14452  
/note="MIR repeat: matches 141. .225 of consensus"  
14589. .14679  
/note="MIR repeat: matches 173. .262 of consensus"  
14597. .15201  
/note="match: GSS: Em:AQ53482"  
14616. .15060  
/note="match: GSS: Em:AQ370601"  
14868. .15040  
/note="MIR repeat: matches 49. .233 of consensus"  
15071. .15188  
/note="L2 repeat: matches 2112. .2239 of consensus"  
15304. .15399  
/note="MLTIB repeat: matches 1. .99 of consensus"  
15490. .15662  
/note="AluSq repeat: matches 2. .114 of consensus"  
15669. .15727  
/note="MLTIB repeat: matches 119. .178 of consensus"  
15728. .16027







RESULT 7  
CEFS7BL/c 36532 bp DNA linear INV 23-JAN-2002  
LOCUS Caenorhabditis elegans cosmid F57BL, complete sequence.  
DEFINITION 278064  
ACCESSION 278064  
VERSION 278064.1 GI:1487912  
KEYWORDS HTG: Collagen.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.  
REFERENCE 1  
AUTHORS none.  
TITLE Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium  
JOURNAL Science 282 (5396), 2012-2018 (1998)  
MEDLINE 99069613  
PUBMED 9851916  
REMARK The C.elegans Sequencing Consortium.  
REFERENCE 2 (bases 1 to 36532)  
AUTHORS Sims,M.A.  
TITLE Direct Submission  
JOURNAL Submitted (09-AUG-1996) Nematode Sequencing Project, Sanger  
Institute, Hinxton, Cambridge CB10 1SA, England and Department of  
Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:  
jes@sanger.ac.uk or tw@nematode.wustl.edu  
Coding sequences below are predicted from computer analysis, using  
predictions from GeneFINDER (P. Green, U. Washington), and other  
available information.  
Current sequence finishing criteria for the C. elegans genome  
sequencing consortium are that all bases are either sequenced  
unambiguously on both strands, or on a single strand with both a  
dye primer and dye terminator reaction, from distinct subclones.  
Exceptions are indicated by an explicit note.  
IMPORTANT: This sequence is not the entire insert of clone F57BL.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we arrange for a small overlap between  
neighbouring submissions.  
The true left end of clone F57BL is at 1 in this sequence. The true  
right end of clone F57BL is at 5273 in  
sequence 278060.  
The true left end of clone C34D1 is at 35789 in this sequence. The  
start of this sequence (1..104) overlaps with the end of sequence  
278065.  
The end of this sequence (35789..36532) overlaps with the start of  
sequence 278060.  
For a graphical representation of this sequence and its analysis  
see: [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=F57BL)  
name=F57BL  
IMPORTANT: This sequence is NOT necessarily the entire insert of  
the specified clone. It may be shorter because we only sequence  
overlapping sections once, or longer because we arrange for a small  
overlap between neighbouring submissions.  
Location/Qualifiers  
1..36532  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/chromosome="V"  
/clone="F57BL"  
complement(join(3851..4162,4210..4629,4674..4877,  
4921..5024,5071..5452))  
/gene="F57BL.2"  
complement(join(3851..4162,4210..4629,4674..4877,  
4921..5024,5071..5452))  
/gene="F57BL.2"  
/note="predicted using GeneFINDER  
cDNA EST YK6699.3 comes from this gene  
cDNA EST YK6699.5 comes from this gene"  
/codon\_start=1  
/protein\_id="CAB01511.1"  
/db\_xref="GI:3877753"  
/db\_xref="SPTREMBL:O20924"

/translation="MALRHTISFQFSNRHSPPYTRSVRGVQPLDSTPYTRRDSO  
PGTITTGTFQRFHESADSEIDLNSKRYFIEHFSYKRTISMKEEMWDLLEYRTLRWR  
RRPVPTWQGRKRLMAVFLPAMLYKYARDCPLFGDTIHNSGSAVDKXANNAEKKQKH  
QTISNLAEFSANDKODKFDLHEKLELVNLSKMKSAIEELKQIKLMOAIEID  
ALOOMKEIDAKSTKTIHSTPEKAPTAPTAIPSSLOPMHITRRALLGVAVS  
LIGASIDHSSSRVSAKDFEFDNMYRECTFOEGVALLDRDLSREANCYDKRAT  
LTVKLARVLPKSVSTOHRVMSGLVPEVHAKLTDVAVACTDSCCKMOPLVANCTYKRR  
DGSVDEQEQFCSVPTI0NHSPINHVQFRFRENHGMKTCAYLIRVYGEPPDKETQ  
PMTDNGTESKLESIAVNSVETA"  
complement(join(5920..6069,6126..6209,6280..6408,  
6453..6566))  
/gene="F57BL.1"  
complement(join(5920..6069,6126..6209,6280..6408,  
6453..6566))  
/gene="F57BL.1"  
/note="predicted using GeneFINDER"  
/codon\_start=1  
/protein\_id="CAB01510.1"  
/db\_xref="GI:3877752"  
/db\_xref="SPTREMBL:O20923"  
/translation="MNNQCYEYDERILMKHYVLSCTSSGIRKNIAPPTIVQTIETILQ  
FILPILIGIYIATIKIILMKKALMKYELTIKQALSSSVFLCOTIFETIATATL  
IKRINTTEICAGAAPCPFFPFSKEIRKLVSRYSLTTSQANSNSQORRQTSRAI"  
complement(join(10251..10976,11025..11240))  
/gene="F57BL.3"  
complement(join(10251..10976,11025..11240))  
/gene="F57BL.3"  
/note="predicted using GeneFINDER  
contains similarity to Pfam domain: PF01391 (collagen  
triple helix repeat (20 copies)). Score=45.4,  
E-value=4e-10, N=3; PF01484 (Nematode cuticle collagen  
N-terminal domain), Score=40.6, E-value=1.1e-08, N=1  
cDNA EST YK7307.5 comes from this gene"  
/codon\_start=1  
/protein\_id="CAB01509.1"  
/db\_xref="GI:3877751"  
/db\_xref="SPTREMBL:O20922"  
/translation="MSSEDPKQIATVETESIRKVAFFGIATVATATLTAIVAPMLINY  
MOHVOSLQSEVEFCAGHRSNGIDVEYRFEGVSGVAGIRIKESYHRKARASKVRSQY  
GADAAVAGFCGASAGSCSCGTGACAGPAGCDGAFNGAGAGAPGNPQDSIEDOT  
AGDPFPCFDPAGPSPGAGGKPGSPGAGVGGTGGALPBPSPGAPGPGSGOPG  
SNCNAGAPGAPGQVGVDPGTPPGAPGPPGAPGAPGQSGSGGPGPGDAGAP  
PGAPGDPGAGAPGQDSGSEBACDHCPPRRAPCY"  
complement(join(13023..13760,13809..13948,14082..14274))  
/gene="F57BL.4"  
complement(join(13023..13760,13809..13948,14082..14274))  
/gene="F57BL.4"  
/note="predicted using GeneFINDER  
contains similarity to Pfam domain: PF01391 (collagen  
triple helix repeat (20 copies)). Score=44.2,  
E-value=9.7e-10, N=3; PF01484 (Nematode cuticle collagen  
N-terminal domain), Score=40.6, E-value=1.1e-08, N=1  
cDNA EST YK9301.3 comes from this gene  
cDNA EST YK70B10.3 comes from this gene  
cDNA EST YK9301.5 comes from this gene  
cDNA EST YK1026.3 comes from this gene  
cDNA EST YK10012.5 comes from this gene  
cDNA EST YK9907.3 comes from this gene  
cDNA EST YK9907.5 comes from this gene  
cDNA EST YK7201.3 comes from this gene  
cDNA EST YK70B10.5 comes from this gene  
cDNA EST YK1026.5 comes from this gene  
cDNA EST YK7201.5 comes from this gene  
cDNA EST YK10017.5 comes from this gene  
cDNA EST YK10017.3 comes from this gene  
cDNA EST YK27906.3 comes from this gene  
cDNA EST YK27906.5 comes from this gene"  
/codon\_start=1  
/protein\_id="CAB01508.1"  
/db\_xref="GI:3877750"  
/db\_xref="SPTREMBL:O20921"  
/translation="MAEHGCVIDKAVGEGKPGNSPEFLASLAPYVYSEYTIASI  
KGMTROWSNHLTRSCPFSHFSTIATLTAIVAPMLKNYQHVOSLSQSEVEFCARHS

NCIMDEYOFNEFEGVSGVAGRIKRESYHRRASGPHRASKYVRQSGYADAAVGGFGSGAG  
 SCCSCGTGAAGPAGSPGODGAPNDGAPCAPENPCODASEDTAGDSCSPCCPAPAP  
 GPGSAGGKPGSPGAPGAGGSGASLPGPPGAPGPGSGGSGNSMAGCAPGAPV  
 DYPGTGAGGPPGPPGAPGAPGPGGSGAGGSGAGGPGGAGGPGGAGGPGGAGGAG  
 DDDSGEGACDCCPPPRAPGT  
 /gene="F57B1.5"  
 /join(19688..19768,19816..19867,19913..20010,20397..20489,  
 20538..20687)  
 /gene="F57B1.5"  
 /join(19688..19768,19816..19867,19913..20010,20397..20489,  
 20538..20687)  
 /gene="F57B1.5"  
 /note="predicted using GeneFinder"  
 /codon\_start=1  
 /protein\_id="CAB01512.1"  
 /db\_xref="GI:3877754"  
 /db\_xref="SPTREMBL:Q20925"  
 /translation="MGENKFKYFVVPFNVLSTKITIILLISLYGMDTKORIS  
 GSNFLAIVDILAIIGASISLWVQKFWKVESFENIGILSLASTISFTFLD  
 PASIOSLTHILILWISAFWAWRETYNNVNLISEWGCGEETIEYLQHF"  
 /join(23062..23112,24099..24151,24322..24499,24547..24696,  
 24748..24830,25045..25267)  
 /gene="F57B1.6"  
 /join(23062..23112,24099..24151,24322..24499,24547..24696,  
 24748..24830,25045..25267)  
 /gene="F57B1.6"  
 /note="predicted using GeneFinder"  
 cDNA EST EMBL:701341 comes from this gene"  
 /codon\_start=1  
 /protein\_id="CAB01513.1"  
 /db\_xref="GI:3877755"  
 /db\_xref="SPTREMBL:Q20926"  
 /translation="MTMNYCGINKLVYDAVVOCKLQSFSAVTLHCETPLGDSRL  
 VSRILIDRKSEPIYKQQLQIVSISGLCTSSVDFYVSALTYTPDHKLINPRM  
 AFWLPPSFTSLISLFLSAGPRVTKFAVDNOMILSLALMCACILSSFS  
 F1ACQANDADIAGYAFATPAQFLASAWYSRLRLVIVSVIOSLHNGIIVSTYLGII  
 SCKRTFYQAEKPPQGLIERTYFA"  
 /complement(join(34555..34692,34753..35192,35241..35370))  
 /gene="F57B1.7"  
 /complement(join(34555..34692,34753..35192,35241..35370))  
 /gene="F57B1.7"  
 /note="predicted using GeneFinder"  
 /codon\_start=1  
 /protein\_id="CAB01507.1"  
 /db\_xref="GI:3877749"  
 /db\_xref="SPTREMBL:Q20920"  
 /translation="WADQIKIRKQKRDTLTPAKNSTSTPSPNSQISLANSILY  
 GSINTSPQLSPSSSDASCSPPMPFPSPSPMPSPADSPPTQPTSPSSIPSSIA  
 STSPMTSLPLRLSGFPLINIRDPALAEASLNLGADNADALAKTILDDYRMLLEASM  
 SMSSSPSKDESDSDGLNSITIDVTDPICLSRAFLVLSNPRPHPLHLTLD  
 CDYSAHCDLPCCPLFS"  
 /complement(join(12157..12228,13962..14312,16459..16593,  
 17111..17131))  
 /gene="Y51H4A.3"  
 /complement(join(12157..12228,13962..14312,16459..16593,  
 17111..17131))  
 /gene="Y51H4A.3"  
 /note="contains similarity to Pfam domain: PF00071 (Ras  
 family), Score=317.2, E-value=6.4e-92, N=1  
 cDNA EST yk60d10.5 comes from this gene  
 cDNA EST yk10347.5 comes from this gene  
 cDNA EST yk77f3.5 comes from this gene

SOURCE  
 ORGANISM  
 Caenorhabditis elegans.  
 Caenorhabditis elegans  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
 Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 REFERENCE  
 AUTHORS  
 TITLE  
 none.  
 JOURNAL  
 MEDLINE  
 PUBMED  
 9851916  
 The C.elegans Sequencing Consortium.  
 REMARK  
 2 (bases 1 to 269619)  
 Sulston, J.E.  
 JOURNAL  
 TITLE  
 Submitted (27-OCT-1999) Nematode Sequencing Project, Sanger  
 Institute, Hinxton, Cambridge CB10 1SA, England and Department of  
 Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:  
 jes@sanger.ac.uk or tw@nematode.wustl.edu  
 On May 14, 2001 this sequence version replaced gi:5730145.  
 Coding sequences below are predicted from computer analysis, using  
 predictions from GeneFinder (P. Green, U. Washington), and other  
 available information.  
 Current sequence finishing criteria for the C. elegans genome  
 sequencing consortium are that all bases are either sequenced  
 unambiguously on both strands, or on a single strand with both a  
 dye primer and dye terminator reaction, from distinct subclones.  
 Exceptions are indicated by an explicit note.  
 For a graphical representation of this sequence and its analysis  
 see: [http://wormbase.sanger.ac.uk/perl/ce/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ce/elegans/seq/sequence?name=y51H4A)  
 name=y51H4A  
 IMPORTANT: This sequence is NOT necessarily the entire insert of  
 the specified clone. It may be shorter because we only sequence  
 overlapping sections once, or longer because we arrange for a small  
 overlap between neighbouring submissions.  
 This sequence is the entire insert of clone Y51H4A. The true left  
 end of clone Y43D4A is at 193825 in this sequence. The start of  
 this sequence (1..115) overlaps with the end of sequence AL713992.  
 The end of this sequence (269514..269619) overlaps with the start  
 of sequence AL13246.  
 Location/Qualifiers  
 1..269619  
 /organism="Caenorhabditis elegans"  
 /db\_xref="taxon:6239"  
 /chromosome="IV"  
 /clone="Y51H4A"  
 /complement(join(2293..2346,2393..2610,2668..2758,  
 3398..3694,5250..5357))  
 /gene="Y51H4A.1"  
 /complement(join(2293..2346,2393..2610,2668..2758,  
 3398..3694,5250..5357))  
 /gene="Y51H4A.1"  
 /note="predicted using GeneFinder"  
 /codon\_start=1  
 /protein\_id="CAB61134.1"  
 /db\_xref="GI:6434492"  
 /db\_xref="SPTREMBL:Q9NAE5"  
 /translation="MLVAVLILFTLILFFYFGKLLIFGKQIVENSIEKINSNIS  
 LAENSKLEKSSQSDNENAEVEYIEHTFAAPENYLEEIOIKVEVEDEDEK  
 PAGSPKSTSCISSEILSROSIGTCAKLCFNTPTPTPIVVRMSSEERSLHKIEK  
 NNCNREFSISPRSPSPDPATSSASLSPNTRPTPIVVRMSSEERSLHKIEK  
 RKSDIDLRLRIQSAKRHRKRSERLQOVQHHHDY"  
 /complement(join(12157..12228,13962..14312,16459..16593,  
 17111..17131))  
 /gene="Y51H4A.3"  
 /complement(join(12157..12228,13962..14312,16459..16593,  
 17111..17131))  
 /gene="Y51H4A.3"  
 /note="contains similarity to Pfam domain: PF00071 (Ras  
 family), Score=317.2, E-value=6.4e-92, N=1  
 cDNA EST yk60d10.5 comes from this gene  
 cDNA EST yk10347.5 comes from this gene  
 cDNA EST yk77f3.5 comes from this gene

```

gene
    cDNA EST yk336a7.5 comes from this gene
    cDNA EST yk435t7.5 comes from this gene
    cDNA EST yk531a12.5 comes from this gene
    cDNA EST yk592a8.5 comes from this gene
    cDNA EST yk651f5.5 comes from this gene
    cDNA EST yk671f3.5 comes from this gene
    cDNA EST yk565h1.5 comes from this gene
    /codon_start=1
    /protein_id="CAB63379.1"
    /db_xref="GI:6580310"
    /translation="MAIRKKIYIVGAGCGKTCILIVSEKDFQPDVVYPIVPEENVYA
    DIEVDSKQVELALMTAGCEDYDRLRLPSTYPTDYLKCFSDSPSLNIEPKKTPPE
    VHFCEPNVPIILVGNKRDLSDPQTVRELAKMKQEPKQEGRAIAEQIGAVLYLCS
    AKTKDIREVFKKATQALQOKKKSKCMIL"
    complement(join(21593..21756,22186..22384,22466..22675))
    /gene="Y51H4A.2"
    CDS
        complement(join(21593..21756,22186..22384,22466..22675))
        /gene="Y51H4A.2"
        /codon_start=1
        /protein_id="CAB63380.1"
        /db_xref="GI:6580311"
        /translation="MKSVGSGSRLLILSRPHRISLPLLGSPSLSEHPDHIY
        IHGGLAAPPOROKALDPKQEHSTATALLCGRVPLLYARKLSEFPHMGAHSS
        VDNKMRKRGRKAKPLITQFISKSFLEPKMKNEVELSTFKSSIFCTFPPKKTFFR
        LIGDIPAPLKLVSRAAEEREREAATRTIV"
        complement(join(31618..31755,32295..32402,33420..33614,
        33718..33834))
        /gene="Y51H4A.4"
        CDS
            complement(join(31618..31755,32295..32402,33420..33614,
            33718..33834))
            /note="predicted using GeneFinder"
            cDNA EST yk282h7.3 comes from this gene
            cDNA EST yk340d3.3 comes from this gene
            cDNA EST yk340d3.3 comes from this gene
            cDNA EST yk479g1.3 comes from this gene
            cDNA EST yk282h7.5 comes from this gene
            cDNA EST yk340d3.5 comes from this gene
            cDNA EST yk479g1.5 comes from this gene
            cDNA EST yk486f5.5 comes from this gene
            cDNA EST yk573a7.3 comes from this gene
            cDNA EST yk573a7.5 comes from this gene
            /codon_start=1
            /protein_id="CAB61136.1"
            /db_xref="GI:6434494"
            /translation="MWTDIQOOLAAASRMCDSISAKFNOMYANPYGAGAFDKNSAA
            TALPFAQHFAPYANPHATSSPSSSTTSSONQSGSSGSKKKRPVYPENOK
            DEAYLDKRRNRNDEAKRSRESKKVDQNSVRVTLERENQCLRTVVOQLQONESMR
            QHLLONPQITIDMPIAMNMPST"
            complement(join(50748..50898,51588..51977,52741..52920,
            53245..53366,53557..53611,54196..54274,54700..54854,
            55009..55055,57258..57575))
            /gene="Y51H4A.5"
            CDS
                complement(join(50748..50898,51588..51977,52741..52920,
                53245..53366,53557..53611,54196..54274,54700..54854,
                55009..55055,57258..57575))
                /note="predicted using GeneFinder
                contains similarity to Pfam domain: PF01764 (lipase),
                Score=244.3, E-value=5.4e-70, N=1"
                /codon_start=1
                /protein_id="CAB61137.1"
                /db_xref="GI:6434495"
                /translation="MERLMCOYLHLFLFYVCWIVGTSALGRAEYTCSHKNTL
                BDATREKSDHDKMKRYVQWCDNTSLYSRQILHSAALRTFIDSAAGSNAITLKL
                LEEDVANNITIKKQKNEKQKRYEENFEKREKLNLLEFFACFNICIPFSNIF
                FNSAANLNTAIIIVFLIFWQFDELEAKCJDCISLGHMCVTVNCGSPCATSIS
                KSLNCELDKAAHYDDNFVRFKIYPLASALPNKECFKPLPLPTMTSKSKSANC
                EYVQSNCSFTSPDQTKVIVMSFRGTSPLOTMEITIDFTFGKQPPDAGNITFY
                FVDAFFFLMNAQLQDIDQLKIKYIPDYELMVTGHSGLGALASVAASTVHTGIFTSDK

```

```

gene
    VKLVTIGQPRTDGYDATWHDKNFYSFRIYHRRDLVAHIIPRODGADELHHRSEWY
    NNMSTTEPHICADEADGFYCSSREDTINADILTYFGVIFEMQTKC"
    join(61826..61986,62058..62127)
    /gene="Y51H4A.6"
    CDS
        join(61826..61986,62058..62127)
        /gene="Y51H4A.6"
        /codon_start=1
        /protein_id="CAB63381.1"
        /db_xref="GI:6580312"
        /translation="MHRNYEKKONLIFSEKQOSVEKYGDDWVHHFTANARETFGK
        ICFIEPRNLSTIERREPRPAKGDLLHNGN"
        complement(join(63978..64270,64412..64705,66455..66576,
        66629..66970,67242..67534,68927..69055,69102..69243,
        69290..69561,70362..70488,70898..71079))
        /gene="Y51H4A.7"
        complement(join(63978..64270,64412..64705,66455..66576,
        66629..66970,67242..67534,68927..69055,69102..69243,
        69290..69561,70362..70488,70898..71079))
        /gene="Y51H4A.7"
        /note="predicted using GeneFinder
        contains similarity to Pfam domain: PF01175 (Urocanase),
        Score=350.6, E-value=5.7e-102, N=1
        cDNA EST yk156d6.5 comes from this gene
        cDNA EST yk156e6.5 comes from this gene
        cDNA EST yk156b6.3 comes from this gene
        cDNA EST yk259b10.3 comes from this gene
        cDNA EST yk259b10.5 comes from this gene"
        /codon_start=1
        /protein_id="CAB61139.1"
        /db_xref="GI:6434497"
        /db_xref="SPTREMBL:O9NAE2"
        /translation="MOSSDFEVVVKRNLPEDEENHONREFGCDLLGKEHTTWEG
        EEKSNIEKSFIDQNNMNIPLVDPEFSLLEHPYEQAKNVAHAKPRPCNTQTEK
        MLAVRNALRIPIKEHNVLLATEFAEELNTGYGYRFPNEDLEFAPVSEIGHCEQ
        ASAIIIMLNNLDKRVAPQPOELVTGNGOVPSNMIOERLVRLYLTMTDHOTLVLY
        SGHPLCFPSTPDSPRMGTIVNGMILPSTKTELYDKYFALGVYQCOMTAGSFCTGPR
        QGIVHSTTTITVLNAGRMGLDSLAGVYPTAGICGSGAOPRAAKIAGCTGVIAEISD
        TALKRHOQGMWDVYSKDLDEIVNMIKEYREKEALISTYLGNAVDMLEERPECL
        VELGSDOTSJLHNFLEGFYPAGLTFEOSQOMSDVKKRKLQNSLIRQIATAIDTA
        AKGMVFDYGNAMFLLECORAANLREDQADDSFVPSYMODIMDDISMGFGPRFM
        VCTSGKPEDIRLTDQACKIIDELKTDQDEYKQYLOLNKKIIEAEKNIKLVGSOA
        RILYSDBRAGRVLAASAFLNLYSGKSAIVLISRHDHVDGSDGSPRETSNVDGSAP
        TADMAVONCIGDSFRCATVYALHNGSGVGGVDVINGFGVILVDSSDAARRABGMLNW
        DVPNGVTRSMWSNNAKQEAIQRAEKQVGLRVTLFVEADEELKRLKF"
        /gene="Y51H4A.1"
        80385..80457
        /gene="Y51H4A.t1"
        80385..80457
        /gene="Y51H4A.t1"
        trNA
            gene
                Query Match 41.0%; Score 33.2; DB 3; Length 269619;
                Best Local Similarity 64.1%; Pred. 3.1;
                Matches 50; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
            Yy 1 GGACCTCAGCACACCCAGAGCCCGAAGAAGTGGGGCTCTCGTGCACACGCTTTA 60
                ||||| ||||| || || ||||| || ||||| || ||||| || ||||| ||
                DB 130146 GGACACACAGACCATCTCGAGCTCGAGACAAAGGACCTTCAGGACACGAGAGCT 130087
                ||||| ||||| || || ||||| || ||||| || ||||| || ||||| ||
                Yy 61 CCACATATACAGAGAGAA 78
                ||| ||| || ||||| ||
                Db 130086 CCAGGACATCTGGAGGA 130069
                ||||| ||||| || || ||||| || ||||| || ||||| || ||||| ||
            RESULT 9
            AF282902 5851 bp mRNA linear INV 17-DEC-2000
            LOCUS AF282902 Hydra vulgaris type IV collagen alpha 1 chain precursor, mRNA,
            DEFINITION complete cds.
            ACCESSION AF282902
            VERSION AF282902.1 GI:11875611
            KEYWORDS
            SOURCE Hydra vulgaris.
            ORGANISM Hydra vulgaris

```









Intron /number-22  
 9721..10695  
 /gene="MUC5B"  
 /number-22  
 10696..10872  
 Exon

Query Match 39.5%; Score 32; DB 9; Length 15143;  
 Best Local Similarity 79.2%; Pred. No. 6.1;  
 Matches 38; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 9 AGCACACCCAGGCCCCAGAGAAGTGGGGCTCTGTGTGACACAG 56  
 || ||||| ||||| || ||||| || || |||||  
 Db 3212 AGGCCACCCAGGACCCAGAGAGGTGGGCTCCGGGGCTGCAAG 3259

RESULT 15  
 HSS171M\_3  
 WPCOMMENT

Sequence split into 4 fragments LOCUS HSS171M Accession AJ239326  
 Fragment Name Begin End  
 HSS171M\_0 1 110000  
 HSS171M\_1 100001 210000  
 HSS171M\_2 200001 310000  
 HSS171M\_3 300001 400516  
 Continuation (4 of 4) of HSS171M from base 300001 (AJ239326 Homo sapiens chromosome 21 c

Query Match 39.5%; Score 32; DB 2; Length 100516;  
 Best Local Similarity 65.3%; Pred. No. 7.1;  
 Matches 47; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 1 GGACCTCAGACACACCCAGGCCCCAGAGAAGTGGGGCTCTGTGTGACACAGTTTA 60  
 || ||||| ||||| || || ||||| || || ||||| || || |||||  
 Db 26459 GGCCCTCCAGAGACCCAGAGGCGCCCAAGAGAGAGTGGGCCCCCGGACCACCAAGGTGAG 26518  
 QY 61 CCACAATATACA 72  
 || || ||||  
 Db 26519 CAACCTGTGACA 26530

Search completed: February 20, 2003, 04:28:59  
 Job time : 431.251 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:48:42 : Search time 1001.14 Seconds  
(without alignments)  
6406.126 Million cell updates/sec

Title: US-09-997-610-1\_COPY\_53\_448

Perfect score: 396  
Sequence: 1 gctggactccagcacacc.....ctgggtgctgttaagca 396

Scoring table:  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estlmu:\*  
5: em\_estlov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_dln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_tod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	113	28.5	499	14	BM967732 LM24HM013
2	76.4	19.3	504	12	BE757275 211668 MA
3	76.4	19.3	526	12	BE757276 211669 MA
4	61.8	15.6	612	17	AG088117 Pan trogl
5	59.6	15.1	538	9	AA777621 2195407.s
6	58.6	14.8	440	10	BE062167 RCI-BT025

Result No.	Score	Query Match	Length	ID	Description
7	58.6	14.8	475	17	B36584 HS-1041-A1-
8	58.6	14.8	537	10	BE079777 RC6-BT062
9	58.6	14.8	563	10	BE079876 RC6-BT062
10	58.6	14.8	678	10	AM813783 RC3-ST019
11	58.6	14.8	727	10	AV731140 AV731140
12	58.6	14.8	2615	10	BE420422 32-393 hu
13	58.6	14.8	345	12	AO012504 CIT-HSP-2
14	58.6	14.8	342	17	BF81529 OVI-ET018
15	58.6	14.8	670	17	AC093990 Pan trogl
16	57.8	14.6	397	13	B1180953 TY4F01 he
17	57.8	14.6	425	10	AM517269 xq08B09.x
18	57.8	14.4	460	9	AA372508 EST84499
19	57.8	14.4	666	17	AC019007 Homo sapi
20	57.8	14.4	666	17	AG141320 Pan trogl
21	57.8	14.4	683	17	AC019080 Homo sapi
22	57.8	14.4	1036	13	BM462674 AGENCOURT
23	56.8	14.3	577	10	AV716883 AV716883
24	56.6	14.3	662	10	BE390081 601285704
25	56.4	14.2	284	9	AA360873 EST70071
26	56.4	14.2	461	12	BF828364 MR1-HN006
27	56.4	14.2	461	12	BF829000 MR1-HN006
28	56.4	14.2	545	10	BE279213 601156706
29	56.4	14.2	609	10	BE389158 601285954
30	56.4	14.2	667	17	AC080755 Pan trogl
31	56.4	14.2	668	17	AC080765 Pan trogl
32	56.4	14.2	698	12	BE729966 601562450
33	56.4	14.2	705	17	AG052122 Pan trogl
34	56.4	14.2	721	12	BE728927 601562364
35	56.4	14.2	761	12	BE728912 601562343
36	56.4	14.2	795	12	BF025976 601669973
37	56.4	14.2	800	12	BF203746 601688725
38	56.4	14.2	885	12	BF304683 601888096
39	56.4	14.2	1080	14	BQ422247 AGENCOURT
40	56.4	14.1	315	9	AA719635 2952604.s
41	56.4	14.1	645	17	AG014562 Homo sapi
42	55.8	14.1	360	10	AM898344 RC3-NN07
43	55.4	14.0	364	17	AO665181 HS.5344.B
44	55.4	14.0	388	17	B78843 CIT-HSP-734
45	55.4	14.0	453	17	AO393565 CITBI-EI-

## ALIGNMENTS

RESULT 1  
LOCUS BM967732 499 bp mRNA linear EST 20-MAR-2002  
DEFINITION LM24HM0134 Bos taurus LM-24-HW cDNA library Bos taurus cDNA clone  
ACCESSION BM967732  
VERSION BM967732.1 GI:19561919  
KEYWORDS EST.  
SOURCE Bos taurus  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 499)  
Yoon,D.H., Jang,Y.S., Kim,T.H., Park,E.W., Lee,H.K., Chung,E.R.,  
Sun,S.S. and Cheong,I.C.  
Gene Expression Profiling of the Bovine skeletal muscle  
Unpublished (2002)  
CONTACT: Dr. Du-Hak Yoon  
National Livestock Research Institute, RDA  
564 Omoekchun-dong, Suwon, 441-350, Korea  
Tel: 82 31 290 1593  
Fax: 82 31 290 1792  
Email: dhyoon@rda.go.kr  
Insert Length: 499 Std Error: 0.00  
Seq primer: CAGGAACACGTATGAC  
POLYA-No.  
FEATURES  
source location/Qualifiers  
1..499

/organism="Bos taurus"									
/db_xref="taxon:9913"									
/clone="LM-24-HW-011-34 (5)"									
/clone_id="Bos taurus LM-24-HW cDNA library"									
/sex="six males mixed"									
/tissue_type="longissimus dorsi"									
/cell_type="myocyte"									
/dev_stage="24 months old"									
/lab_host="XLI-BlueMFL strain"									
/note="Organ: skeletal muscle; Vector: Uni-ZAPXR; Site:1; EcorI; Site_2: Xho I"									
BASE COUNT	129 a	139 c	131 g	100 t					
ORIGIN									
Query Match	28.5%	Score 113;	DB 14;	Length 499;					
Best Local Similarity	64.7%;	Pred. No. 5.8e-21;							
Matches 200;	Conservative 0;	Mismatches 105;	Indels 4;	Gaps 2;					
QY	2	CTGACCTTCACACACACCCACAGGCCCGCAGAAAGTGGGGCCCTCGTGGTGACACAGTT	61						
Db	181	CGGGCCACACAGACACCCAGGCGCTCCAGGTATAGAGGGCCCTCAGTATAAAGCGAA	240						
QY	62	TACCAAAATATACAGGA--GAATTAAGTGAATGACAAATGCCCTGCTGATATAG	118						
Db	241	TACACAGTTTGGCAGAGTCGCCGGGAGACTCCAGACCAAGTGTAATAGCCATGCCACA	300						
QY	119	AAAGTTCAGCCTTTACTGTGAAGCTCAGTGCAGAAACCTTCCTCTTCAAGCCATCA	178						
Db	301	GACACTCCGCGCTTCACTGTGAAGCTCAGTGGCGACATTCCTCCCTCAAGCCTGTGC	360						
QY	179	TC-TTCACAGGGGCTCTGTACAAATGCCAGAGAGATTTAAAGAGAGCCATGGACCTTT	237						
Db	361	CCTTTCACAGAGTCTCTGTACATGCCAGAAAGCTTACAGAGAGACCTGGGCTTTC	420						
QY	238	GCTTGCAGGGTCCCTGGGAATTTACTCTCAGCTTTGATGTTGAGCTGATCATTTGCAAG	297						
Db	421	ACATGCAGGGTCCAGAAATTACCATTTCCCTTTCTATGTGATCTCATCATCAGCAAG	480						
QY	298	GTGAATATT 306							
Db	481	GTGACTGTT 489							
RESULT 2									
LOCUS	BE757275	504 bp	mRNA	linear	EST 25-APR-2001				
DEFINITION	211668 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.								
ACCESSION	BE757275								
VERSION	BE757275.1	GI:10171267							
KEYWORDS	EST.								
SOURCE	COW.								
ORGANISM	Bos taurus								
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;								
	Bovidae; Bovinae; Bos.								
REFERENCE	1 (bases 1 to 504)								
AUTHORS	Smith,T.P.L., Grose,W.M., Feking,B.A., Roberts,A.J., Stone,R.T.,								
	Caas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett								
	G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,								
	Petee,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and								
	Keefe,J.W.								
TITLE	Sequence evaluation of four pooled-tissue normalized bovine cDNA								
JOURNAL	libraries and construction of a gene index for cattle								
MEDLINE	Genome Res. 11 (4), 626-630 (2001)								
COMMENT	21180013								
	Contact: Smith TPL								
	USDA, ARS, US Meat Animal Research Center								
	PO Box 166, Clay Center, NE 68933-0166, USA								
	Tel.: 402 762 4366								
	Fax: 402 762 4390								
	Email: smithth@mail.marc.usda.gov								
	Single pass sequencing. Bases called and alt_trimmed with phred								
	V0.180904.e. Vector identified by cross_match with the -mismatch 18								

	and -mismatch 12 options.					
	PCR Primers					
	FORWARD: AGAACAACAGCTATGACCAT					
	BACKWARD: GTTTTCCAGTCACAGC					
	Plate: 63 row: G column: 13					
	Seq primer: ATTAGGTGACACTATGA.					
FEATURES	Location/Qualifiers					
source	1..504					
	/organism="Bos taurus"					
	/db_xref="taxon:9913"					
	/clone_lib="MARC 2BOV"					
	/tissue_type="Pooled"					
	/lab_host="DH10B"					
	/note="vector: pCMV SPORT6; site_1: NotI; site_2: SalI;					
	library made from pooled tissue from testis, thymus,					
	semitendinosus muscle, longissimus muscle, pancreas,					
	adrenal, and endometrium."					
BASE COUNT	99 a	131 c	146 g	128 t		
ORIGIN						
Query Match	19.3%	Score 76.4;	DB 12;	Length 504;		
Best Local Similarity	61.6%;	Pred. No. 1.1e-10;				
Matches 122;	Conservative 0;	Mismatches 76;	Indels 0;	Gaps 0;		
OY	123	GTGCACCTTACTGTGAACTCGAGGGAATACTCTCTCTTCAAGGCCATCATCTT	182			
Db	303	GTCTCCCTTTTCCGTACAGCTGAGGGCCCTTCCAGAGACCTCCAGCCATTGTCTT	362			
OY	183	CACAGGGGTCTGTACATGATGCCAGAGGATTAAAGAGGACATGAGTGCTTGTG	242			
Db	363	CAGAAGTGTCTGTACAAOCATCAGGGCCACTTGCACCCGCACTGGTGTTCAGCG	422			
OY	243	CAGGCTGCCGTGGAAATTACTACTCCAGCTTGATGTGAGCTGCATCATTCGAGGTGA	302			
Db	423	CAGCGTCCCTGGTGTACACACTTTGGCTTGCATTGAGTTGTTTCAGAGTGTGCAA	482			
OY	303	TATTTGGCTAATGAGGA 320				
Db	483	GGTGGGTCTAATGCGGA 500				
RESULT 3						
BE757276						
LOCUS	BE757276	526 bp	mRNA	linear	EST 25-Apr-2001	
DEFINITION	211669 MARC 2BOV Bos taurus CDNA 5', mRNA sequence.					
ACCESSION	BE757276					
VERSION	BE757276.1 GI:10171268					
KEYWORDS	EST.					
SOURCE	cov.					
ORGANISM	Bos taurus					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;					
	Bovidae; Bovinae; Bos.					
	1 (bases 1 to 526)					
REFERENCE	Smith,T.P.L., Grose,M.M., Freking,B.A., Roberts,A.J., Stone,R.T.,					
AUTHORS	Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett					
	,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,					
	Petee,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J., and					
	Keele,J.W.					
TITLE	Sequence evaluation of four pooled-tissue normalized bovine CDNA					
JOURNAL	libraries and construction of a gene index for cattle					
MEDLINE	Genome Res. 11 (4), 625-630 (2001)					
COMMENT	21180013					
	Contact: Smith TPL					
	USDA, ARS, US Meat Animal Research Center					
	PO Box 166, Clay Center, NE 68933-0166, USA					
	Tel: 402 762 4366					
	Fax: 402 762 4390					
	Email: smith@email.marc.usda.gov					
	Single pass sequencing. Bases called and alt trimmed with phred					
	v0.980904.e. Vector identified by cross_match with the -minscore 18					
	and -mismatch 12 options.					
	PCR primers					





ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 537)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-RC6-BR0627-140  
200-011-A05&f3=2000-02-14&tl=1)  
Seq primer: puc 18 forward  
High quality sequence start: 69  
High quality sequence stop: 537.  
Location/Qualifiers  
1. .537  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BR0627"  
/dev\_stage="Adult"  
/note="Organ: breast; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
BASE COUNT 163 a 135 c 100 g 137 t 2 others  
ORIGIN  
Query Match 14.8%; Score 58.6; DB 10; Length 537;  
Best Local Similarity 75.3%; Pred. No. 1.1e-05;  
Matches 73; Conservative 0; Mismatches 24; Indels 0; Gaps 0;  
QY 300 GAATATTGGCTAATGAGAACCAATTGGCTAATAAGAGAAATTTCTAACGACA 359  
DB 315 GAACTTTGAAGCTTCAGAGAGATGATTAAAGTGTGCTGAGAAATTTCTAACGACA 256  
QY 360 AAGCATTCAGAGAGTGAAGTGGTCTGTTAAAGCA 396  
DB 255 AAGCATTCAGAGAGTGAAGTGGTCTGTTAAAGCA 219  
RESULT 9  
LOCUS BE079876/c 563 bp mRNA linear EST 12-JUN-2000  
DEFINITION RC6-BR0627-220300-012-H08 BR0627 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BE079876  
VERSION BE079876.1 GI:8470160  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 563)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-RC6-BR0627-220  
300-012-H08&f3=2000-03-22&tl=1)  
Seq primer: puc 18 forward  
High quality sequence start: 32  
High quality sequence stop: 563.  
Location/Qualifiers  
1. .563  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BR0627"  
/dev\_stage="Adult"  
/note="Organ: breast; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
BASE COUNT 174 a 144 c 100 g 144 t 1 others  
ORIGIN  
Query Match 14.8%; Score 58.6; DB 10; Length 563;  
Best Local Similarity 75.3%; Pred. No. 1.1e-05;  
Matches 73; Conservative 0; Mismatches 24; Indels 0; Gaps 0;  
QY 300 GAATATTGGCTAATGAGAACCAATTGGCTAATAAGAGAAATTTCTAACGACA 359  
DB 341 GAACTTTGAAGCTTCAGAGAGATGATTAAAGTGTGCTGAGAAATTTCTAACGACA 282  
QY 360 AAGCATTCAGAGAGTGAAGTGGTCTGTTAAAGCA 396  
DB 281 AAGCATTCAGAGAGTGAAGTGGTCTGTTAAAGCA 245  
RESULT 10  
LOCUS AM813783/c 678 bp mRNA linear EST 17-MAY-2000  
DEFINITION RC3-ST0197-120200-015-a03 ST0197 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AM813783  
VERSION AM813783.1 GI:7906777  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 678)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663

[illegible]

Db 553 AAGCATTCAGAGTAACCTGGTGCTGTTAAAGCA 589

RESULT 13

LOCUS A0012504/c 342 bp DNA linear GSS 06-JUN-1998

DEFINITION CIT-HSP-2298M4.TP CIT-HSP Homo sapiens genomic clone 2298M4, DNA sequence.

ACCESSION A0012504

VERSION A0012504.1 GI:3185069

KEYWORDS GSS.

ORGANISM Homo sapiens

SOURCE human.

REFERENCE 1 (bases 1 to 342)

AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.

Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)

TITLE Unpublished (1998)

JOURNAL Contact: Mark Adams

COMMENT Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC end search page:  
[http://www.tigr.org/tdb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html).  
Seq primer: M13-21  
Class: BAC ends.

FEATURES

source Location/Qualifiers

1..342

/organism="Homo sapiens"

/db\_xref="GDB:7153856"

/db\_xref="taxon:9606"

/clone="2298M4"

/clone\_lib="CIT-HSP"

/sex="Male"

/cell\_type="Sperm"

/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2: HindIII"

BASE COUNT 88 a 85 c 58 g 111 t

ORIGIN

Query Match 14.6%; Score 58; DB 17; Length 342;  
Best Local Similarity 100.0%; Pred. No. 1.4e-05;  
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 339 GGAAGAAATTTCTAAGCAGCAAGATTCAGAGTGGTGCTGTTAAAGCA 396  
|||||  
Db 264 GGAAGAAATTTCTAAGCAGCAAGATTCAGAGTGGTGCTGTTAAAGCA 207

RESULT 14

LOCUS BF881529/c 345 bp mRNA linear EST 17-JAN-2001

DEFINITION OVI-ET0181-031200-546-f03 ET0181 Homo sapiens cDNA, mRNA sequence.

ACCESSION BF881529

VERSION BF881529.1 GI:12271655

KEYWORDS EST.

ORGANISM Homo sapiens

SOURCE human.

REFERENCE 1 (bases 1 to 345)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV1&tc2=QV1-ET0181-031200-546-f03&tc3=2000-12-03&tl=1>)  
Seq primer: puc 18 forward  
High quality sequence stop: 89.

FEATURES

source Location/Qualifiers

1..345

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="ET0181"

/dev\_stage="Adult"

/note="Organ: Lung-tumor; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 86 a 96 c 57 g 106 t

ORIGIN

Query Match 14.6%; Score 58; DB 12; Length 345;  
Best Local Similarity 100.0%; Pred. No. 1.4e-05;  
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 339 GGAAGAAATTTCTAAGCAGCAAGATTCAGAGTGGTGCTGTTAAAGCA 396  
|||||  
Db 194 GGAAGAAATTTCTAAGCAGCAAGATTCAGAGTGGTGCTGTTAAAGCA 137

RESULT 15

LOCUS AG093990/c 670 bp DNA linear GSS 03-NOV-2001

DEFINITION Pan troglodytes DNA, clone: PTB-094J23.R, genomic survey sequence.

ACCESSION AG093990

VERSION AG093990.1 GI:16645792

KEYWORDS GSS.

ORGANISM Pan troglodytes

SOURCE BAC library clone:PTB-094J23.R.

REFERENCE 1

AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE BAC end sequences of library PTB

JOURNAL Unpublished

COMMENT 2 (bases 1 to 670)  
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan 1-7-22 Suenho-chou,Tsukuba-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpes@gsc.riken.go.jp, URL:<http://hgp.gsc.riken.go.jp/>, Tel:81-45-503-9111, Fax:81-45-503-9170)  
Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the Rad process and may have higher chance of

clone tracking errors.

PRIMERS

Sequencing: M13Rev

LIBRARY

Vector : PKS145

R.Site 1 : SacI

R.Site 2 : SacI.

Location/Qualifiers

1. .670

/organism="Pan troglodytes"

/db\_xref="taxon:9598"

/clone="PTB-094J23.R"

/sex="male"

/cell\_type="lymphoblast"

/clone\_lib="PTB Chimpanzee Male BAC Library"

BASE COUNT 155 a 186 c 124 g 205 t

ORIGIN

Query Match 14.6%; Score 58; DB 17; Length 670;

Best Local Similarity 100.0%; Pred.No. 1.7e-05;

Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 339 GGAGAAATTTCTAAGCAGCAAGCATTCAGAGTGACTGGTGTGTTAAAGCA 396

DB 404 GGAGAAATTTCTAAGCAGCAAGCATTCAGAGTGACTGGTGTGTTAAAGCA 347

Search completed: February 20, 2003, 06:14:33

Job time : 1009.14 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 18:01:21 : Search time 31.8656 Seconds  
(without alignments)  
6329.459 Million cell updates/sec

Title: US-09-997-610-1\_COPY\_53\_448

Perfect score: 396

Sequence: 1 gctggacctccagcacacc.....cttggtgctgttaagca 396

Scoring table:

IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 424239 seqs, 25461826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published\_Applications\_NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCRT\_NEM\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEM\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEM\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCRTS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEM\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEM\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEM\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEM\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being predicted, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	396	100.0	1381	9	US-09-997-610-1
2	396	100.0	1731	9	US-09-997-610-5
3	288	72.7	1377	9	US-09-997-610-3
4	288	72.7	1731	9	US-09-997-610-7
5	114.8	29.0	425	10	US-09-960-352-15057
6	93	23.5	415	10	US-09-960-352-9137
7	91.6	23.1	389	10	US-09-960-352-4220
8	82.2	20.8	392	10	US-09-960-352-1786
9	78.4	19.8	410	10	US-09-960-352-1715
10	76	19.2	374	10	US-09-960-352-11516
11	73.2	18.5	447	10	US-09-960-352-3684
12	58.8	14.8	467	10	US-09-960-352-4241
13	58.4	14.7	273	10	US-09-864-761-25080
14	58.4	14.7	519	10	US-09-864-761-8349
15	57	14.4	1946	10	US-09-864-761-2925
16	57	14.4	2503	9	US-09-819-607-3
17	56	14.1	505	10	US-09-864-761-7075
18	55.4	14.0	520	10	US-09-864-761-8643
19	55.4	14.0	542	10	US-09-864-761-8938

20	55.4	14.0	170834	10	US-09-835-232-7	Sequence 7, Appl1
21	55.4	14.0	170834	10	US-09-835-232-7	Sequence 7, Appl1
22	54.8	13.8	525	10	US-09-864-761-8932	Sequence 8932, Ap
23	53.8	13.6	444	10	US-09-864-761-24780	Sequence 24780, A
24	53.8	13.6	444	10	US-09-864-761-83	Sequence 83, Appl
25	53.8	13.6	523	10	US-09-864-761-8787	Sequence 8787, Ap
26	53.8	13.6	543	10	US-09-864-761-8034	Sequence 8034, Ap
27	53.8	13.6	544	10	US-09-864-761-9454	Sequence 9454, Ap
28	53.2	13.4	531	10	US-09-864-761-14775	Sequence 14775, A
29	53.2	13.4	576	10	US-09-864-761-8294	Sequence 9294, Ap
30	53.2	13.4	580	10	US-09-864-761-9802	Sequence 9802, Ap
31	52.6	13.3	552	10	US-09-864-761-8495	Sequence 8495, Ap
32	52.2	13.2	220	10	US-09-864-761-33019	Sequence 33019, A
33	52.2	13.2	516	10	US-09-864-761-7399	Sequence 7399, Ap
34	52.2	13.2	520	10	US-09-864-761-8449	Sequence 8449, Ap
35	52.2	13.2	571	10	US-09-864-761-16494	Sequence 16494, A
36	52.2	13.2	600	10	US-09-864-761-7441	Sequence 7441, Ap
37	52.2	13.2	202001	10	US-09-734-674-3	Sequence 3, Appl1
38	51.6	13.0	473	10	US-09-864-761-2556	Sequence 2556, Ap
39	51.6	13.0	3816	10	US-09-880-107-2298	Sequence 2298, Ap
40	51.6	13.0	42899	10	US-09-740-029-3	Sequence 3, Appl1
41	51.6	13.0	155074	9	US-10-026-188-6	Sequence 6, Appl1
42	51.6	13.0	180557	12	US-10-003-806-6	Sequence 6, Appl1
43	51.6	13.0	180557	12	US-10-003-806-9	Sequence 9, Appl1
44	51.2	12.9	148567	9	US-10-254-869-3	Sequence 3, Appl1
45	51.2	12.9	148567	10	US-09-801-876B-3	Sequence 3, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-997-610-1  
Sequence 1, Application US/09997610  
Patent No. US20020156244A1  
GENERAL INFORMATION:  
APPLICANT: Fox, Brian  
APPLICANT: Holloway, James L.  
TITLE OF INVENTION: ADIPOCYTE COMPLEMENT RELATED PROTEIN  
FILE REFERENCE: ZACRP13  
CURRENT APPLICATION NUMBER: US/09/997,610  
CURRENT FILING DATE: 2001-11-29  
PRIOR APPLICATION NUMBER: US 60/253,924  
PRIOR FILING DATE: 2000-11-29  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 1381  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (2)...(1381)  
US-09-997-610-1  
Query Match 100.0%; Score 396; DB 9; Length 1381;  
Best Local Similarity 100.0%; Pred. No. 3.3e-114;  
Matches 396; Conservative 0; Mismatch 0; Indels 0; Gaps 0;  
1 GCTGGACCTCCAGCACACCCAGCCGCCAGGAAGTGGGCTCTCTGTCACCAAGT 60  
|||||  
53 GCTGGACCTCCAGCACACCCAGCCGCCAGGAAGTGGGCTCTCTGTCACCAAGT 112  
|||||  
61 TTACCAATATATACAGAGAAATTAAGTGAATGCAAAATGCCCTGCTCTATAGAA 120  
|||||  
113 TTACCAATATATACAGAGAAATTAAGTGAATGCAAAATGCCCTGCTCTATAGAA 172  
|||||  
121 AGGTGACCTTTACTGTGAAGCTCACTGGAATGCAAACTCTCTTCAAGCCATATC 180  
|||||  
173 AGGTGACCTTTACTGTGAAGCTCACTGGAATGCAAACTCTCTTCAAGCCATATC 232  
|||||  
181 TTACAGGGGTCTGTATACATGCCAGAGGAGTTTAAAGAGCCATGGAGTCTTGCT 240

```
Db 233 TTCACAGGGGTCCTGTACATATGCCAGAGGAGATTAAAGAGGCCATGGAGCTTTGCT 292
Qy 241 TGCAGGGTGCCTGGGAATTACTACAGCTTTGATGTTGAGTGCATCTTCAAGGT 300
Db 293 TGCAGGGTGCCTGGGAATTACTACAGCTTTGATGTTGAGTGCATCTTCAAGGT 352
Qy 301 AATATTTGGCTAATGAGGAACAATTTGGCTAATTAAGAGAAATTTCTTAAGACGAA 360
Db 353 AATATTTGGCTAATGAGGAACAATTTGGCTAATTAAGAGAAATTTCTTAAGACGAA 412
Qy 361 AGCATTCAGAGGAGTGCCTGGCTGCCTTAAGGCA 396
Db 413 AGCATTCAGAGGAGTGCCTGGCTGCCTTAAGGCA 448
```

```
RESULT 2
US-09-997-610-5
; Sequence 5, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLEMENT RELATED PROTEIN
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 1731
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1731)
US-09-997-610-5
```

Query Match 100.0%; Score 396; DB 9; Length 1731;  
Best Local Similarity 100.0%; Pred. No. 3.7e-114; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 GCTGAGCTTCACAGACACCCAGCCGCCAGAGAAGTGGGGCTCCTGTCACCAAGT 60
Db 406 GCTGAGCTTCACAGACACCCAGCCGCCAGAGAAGTGGGGCTCCTGTCACCAAGT 465
Qy 61 TTACCAATATATACAGAGAAATAGTGAATGACAAATGCCCTGTCTGTATATGAA 120
Db 466 TTACCAATATATACAGAGAAATAGTGAATGACAAATGCCCTGTCTGTATATGAA 525
Qy 121 AGGTACAGCTTTACTGTCGAGTCACTGGAATCTCTCTCTTTCAAGCCCATCATC 180
Db 526 AGGTACAGCTTTACTGTCGAGTCACTGGAATCTCTCTCTTTCAAGCCCATCATC 585
Qy 181 TTACAGAGGGTCTCTGTACATGCCAGAGAGATTAAAGAGGCCATGGAGTCTTGGT 240
Db 586 TTACAGAGGGTCTCTGTACATGCCAGAGAGATTAAAGAGGCCATGGAGTCTTGGT 645
Qy 241 TGCAGGGTGCCTGGGAATTACTACTCAGCTTTGATGTTGAGTGCATCTTCAAGGT 300
Db 646 TGCAGGGTGCCTGGGAATTACTACTCAGCTTTGATGTTGAGTGCATCTTCAAGGT 705
Qy 301 AATATTTGGCTAATGAGGAACAATTTGGCTAATTAAGAGAAATTTCTTAAGACGAA 360
Db 706 AATATTTGGCTAATGAGGAACAATTTGGCTAATTAAGAGAAATTTCTTAAGACGAA 765
Qy 361 AGCATTCAGAGGAGTGCCTGGCTGCCTTAAGGCA 396
Db 766 AGCATTCAGAGGAGTGCCTGGCTGCCTTAAGGCA 801
```

```
RESULT 3
US-09-997-610-3
; Sequence 3, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLEMENT RELATED PROTEIN
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1377
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide encoding a polypeptide
; NAME/KEY: variation
; LOCATION: (1)...(1377)
; OTHER INFORMATION: Each n is independently A, T, G, or C.
; NAME/KEY: misc.feature
; LOCATION: (1)...(1377)
; OTHER INFORMATION: n = A,T,C or G
US-09-997-610-3
```

Query Match 72.7%; Score 288; DB 9; Length 1377;  
Best Local Similarity 60.5%; Pred. No. 2.3e-80; Mismatches 239; Conservative 87; Indels 69; Gaps 0;

```
Qy 1 GCTGAGCTTCACAGACACCCAGCCGCCAGAGAAGTGGGGCTCCTGTCACCAAGT 60
Db 52 GCGAGCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNC 111
Qy 61 TTACCAATATATACAGAGAAATAGTGAATGACAAATGCCCTGTCTGTATATGAA 120
Db 112 TTACCAATATATACAGAGAAATAGTGAATGACAAATGCCCTGTCTGTATATGAA 171
Qy 121 AGGTACAGCTTTACTGTCGAGTCACTGGAATCTCTCTCTTTCAAGCCCATCATC 180
Db 172 GGTACAGCTTTACTGTCGAGTCACTGGAATCTCTCTCTTTCAAGCCCATCATC 231
Qy 181 TTACAGAGGGTCTCTGTACATGCCAGAGGATTTAAAGAGGCCATGGAGTCTTGGT 240
Db 232 TTACAGAGGGTCTCTGTACATGCCAGAGGATTTAAAGAGGCCATGGAGTCTTGGT 291
Qy 241 TGCAGGGTGCCTGGGAATTACTACTCAGCTTTGATGTTGAGTGCATCTTCAAGGT 300
Db 292 TGCAGGGTGCCTGGGAATTACTACTCAGCTTTGATGTTGAGTGCATCTTCAAGGT 351
Qy 301 AATATTTGGCTAATGAGGAACAATTTGGCTAATTAAGAGAAATTTCTTAAGACGAA 360
Db 352 AATATTTGGCTAATGAGGAACAATTTGGCTAATTAAGAGAAATTTCTTAAGACGAA 411
Qy 361 AGCATTCAGAGGAGTGCCTGGCTGCCTTAAGGCA 395
Db 412 WSNATTCARAGRGATGCTGGTGTCTTAAGGCA 446
```

```
RESULT 4
US-09-997-610-7
; Sequence 7, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLEMENT RELATED PROTEIN
```



```
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 220
; LENGTH: 389
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (277)(335)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 01-LIB34-084-01-E1-A9
US-09-960-352-220
```

Query Match 23.1%; Score 91.6; DB 10; Length 389;

Best Local Similarity 76.7%; Pred. No. 4.4e-19; Indels 0; Gaps 0;

Matches 112; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
OY 183 CACAGGGGTCCTGTACATCCCGAGAGGATTTAAAGAGGCGCATCTTGGCTTG 242
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 CCCACGGGTCAGTACATGCCAGAGAGACTTACAGAGAGACACTGGGCTTTCACATG 60
OY 243 CAGGTCGCTGGGATTTACTACTCCAGCTTGTGATGCTGCATCATTCGAAGTGAA 302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 CAGGGTCGCCGGAATATACCATTTCTTGTGATGTGATCTCATCATCTGCAAGTGAC 120
OY 303 TATTTGGCTAATGAGAGCAAAATT 328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 TGTTCAGCTGATGAGGACAAAGTT 146
```

## RESULT 8

```
US-09-960-352-1786
; Sequence 1786, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 1786
; LENGTH: 392
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 08-LIB34-023-Q1-E1-B7
US-09-960-352-1786
```

Query Match 20.8%; Score 82.2; DB 10; Length 392;

Best Local Similarity 61.9%; Pred. No. 3.9e-16; Indels 3; Gaps 1;

Matches 148; Conservative 0; Mismatches 88; Indels 3; Gaps 1;

```
OY 2 CTGAGCCTCCAGACACCCAGGCCCCAGAGAGAGTGGGCTCTCGTGACACAGCTT 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 153 CGGGGCGCCAGGACACCCAGGCGCTCCAGGTATACAGAGGCGCTCCGAGTAAAGAGAA 212
OY 62 TACCACAAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 118
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 213 TACCAGGTTTGCAGAGTCCGCGGAGAACTCCAGAGACCAAGTAAATGCCCATGCCACA 272
```

```
OY 119 AAGGTACGCTTTACTGTGAAGCTCAGTGGAAACCTCTCTCTTTCAGACCCATCA 178
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 273 GACAGTCGCGCTTCACGTGATGAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 332
OY 179 TCTTCACAGGAGGCTTCTTACAAATGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 333 CTTTCACAGAGAGGCTTCTTACAAATGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 391
```

## RESULT 9

```
US-09-960-352-3715
; Sequence 3715, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 3715
; LENGTH: 410
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (378)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 16-LIB34-049-01-E1-D11
US-09-960-352-3715
```

Query Match 19.8%; Score 78.4; DB 10; Length 410;

Best Local Similarity 54.9%; Pred. No. 6.3e-15; Indels 6; Gaps 1;

Matches 179; Conservative 0; Mismatches 141; Indels 6; Gaps 1;

```
OY 2 CTGAGCCTCCAGACACCCAGGCCCCAGAGAGAGTGGGCTCTCGTGACACAGCTT 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 18 CAGGCGCTCCCGGCTTCTGCGCTCTGCGCTCTGCGCTCCAGAGGCGCCCGGCTTCAGAGAT 77
OY 62 TACCACAAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78 TCGAGAGACCAACTGGGATTCGCGGAGACATTTGAGAGTGTCTGCTCCCTTAA---- 133
OY 122 GGTACGCTTTACTGTGAAGCTCAGTGAAGAACTCTCTCTCTTTCAGAGCCATCATCT 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134 -TCTGCTTTTGCAGTAAAGATGATGATCCCTCCAGGCGCCCTCCCAACCCATTTGCT 191
OY 182 TCACAGGCGCTCTGTACATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 192 TCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 251
OY 242 CAGAGGTCCTGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 252 GCACACATCTCTGCGGTATACCGCTTGTGCTTGTGATGATGATGATGATGATGATGAT 311
OY 302 ATATTTGGCTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 312 AGCTAGGCTCAGTAAAGATGAGACT 337
```

## RESULT 10

```
US-09-960-352-11516
; Sequence 11516, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
```

APPLICANT: Mathialagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE OF INVENTION: MUSCLE AND FAT DEPOSITION  
FILE REFERENCE: 16511.006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960,352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 11516  
LENGTH: 374  
TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 49-LIB34-046-Q1-E1-E2  
US-09-960-352-11516

Query Match 19.2% Score 76; DB 10; Length 374;  
Best Local Similarity 60.8%; Pred. No. 3.4e-14;

Matches 124; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 123 GTGACGCTTACTGTGAGAGTGAAGAACTCTCTCTTTCAGCCCATCATCTT 182  
DB 171 GTCTGCTTTTCGCTGACCTAGTGGGCTTTCCAGAGACCTCCCATTTGCTT 230  
QY 183 CACAGGGCTCTGTACATGCCAGAGGATTTAAAGAGCCCATGGAGTCTTGTG 242  
DB 231 CCAGGAAGCTCTGTACACCATCAGGCGCCACTCGACCCGCACTGTGTCAACTG 290  
QY 243 CAGGGCTCTGGGAATTAATCTCCAGCTTGTGAGTGCATCATTCAGAGTGAA 302  
DB 291 CAGCGTCCCTGGTGTGTACACTTGGCTTGTGACATTTAGTGTTCAGAGTCTGTCAA 350  
QY 303 TATTTGGCTAATGAGAGCAAT 326  
DB 351 AGGTGCTTAATGCGGAATGAAAT 374

RESULT 11  
US-09-960-352-3684  
Sequence 3684, Application US/09960352  
Patent No. US20020137139A1

GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengbing  
APPLICANT: Byatt, John C.  
APPLICANT: Mathialagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE OF INVENTION: MUSCLE AND FAT DEPOSITION  
FILE REFERENCE: 16511.006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960,352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 3684  
LENGTH: 447  
TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 16-LIB34-024-Q1-E1-D7  
US-09-960-352-3684

Query Match 18.5% Score 73.2; DB 10; Length 447;  
Best Local Similarity 60.6%; Pred. No. 2.8e-13;

Matches 120; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 123 GTGACGCTTACTGTGAGAGTGAAGAACTCTCTCTTTCAGCCCATCATCTT 182  
DB 160 GTCTGCTTTTCGCTGACCTAGTGGGCTTTCCAGAGACCTCCCATTTGCTT 219  
QY 183 CACAGGGCTCTGTACATGCCAGAGGATTTAAAGAGCCCATGGAGTCTTGTG 242  
DB 220 CCAGGAAGCTCTGTACACCATCAGGCGCCACTCGACCCGCACTGTGTCAACTG 279  
QY 243 CAGGGCTCTGGGAATTAATCTCCAGCTTGTGAGTGCATCATTCAGAGTGAA 302  
DB 280 CAGCGTCCCTGGTGTGTACACTTGGCTTGTGACATTTAGTGTTCAGAGTCTGTCAA 339

QY 303 TATTTGGCTAATGAGAGCA 320  
DB 340 GGTGGCTTAATGCGGA 357

RESULT 12  
US-09-960-352-4241  
Sequence 4241, Application US/09960352  
Patent No. US20020137139A1

GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengbing  
APPLICANT: Byatt, John C.  
APPLICANT: Mathialagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE OF INVENTION: MUSCLE AND FAT DEPOSITION  
FILE REFERENCE: 16511.006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960,352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 4241  
LENGTH: 467  
TYPE: DNA  
ORGANISM: Bos taurus  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (377)  
OTHER INFORMATION: unsure at all n locations  
OTHER INFORMATION: Clone ID: 18-LIB34-077-Q1-E1-E5  
US-09-960-352-4241

Query Match 14.8% Score 58.8; DB 10; Length 467;  
Best Local Similarity 54.9%; Pred. No. 9.3e-09;  
Matches 139; Conservative 0; Mismatches 108; Indels 6; Gaps 1;

QY 2 CTGACCTCCAGACACACCCAGGCGCCCAAGAAAGTGGGCTCTGTGACAGGTT 61  
DB 217 CAGGCGCTCCGCGCTCTCTGCGCTCTGTGCGCCCTCCAGAGGCCCAAGGCTCCAGAGT 276  
QY 62 TACCACATATATACAGAGAGAAATAGTAATGACAAATGCCCCCTCTGATATAGAAA 121  
DB 277 TGGGAGGACCACTGGAGATTCGCGAGACATTTAGAGTTGCTGTCTCCCTAA-- 332  
QY 122 GGTACCTTTACTGTGAAGCTAGTGAAGAAATCTCTCTCTTCAAGCCCATCATCT 181  
DB 333 -TCTGCTTTGAGTGAAGATGAATATGCTCCAGGCGCCCTCCAGGCGCCCTCAAGCTGTCT 390  
QY 182 TCACAGGGGCTGTACATGATGAGGAGGATTTAAAGAGAGGAGGAGTGTGCTT 241  
DB 391 TCAGAGGAACCTGCAATTAATGACAGGACCATTTCAATTTACAGGAGTGTCACT 450  
QY 242 GCAGGCTGCTGG 254  
DB 451 GCACCATCCCTGG 463

RESULT 13  
US-09-864-761-25080  
Sequence 25080, Application US/09864761  
Patent No. US20020048763A1

GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aeonica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456

```

? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: US 09/632,366
? PRIOR FILING DATE: 2000-08-03
? PRIOR APPLICATION NUMBER: GB 24263.6
? PRIOR FILING DATE: 2000-10-04
? PRIOR APPLICATION NUMBER: US 60/236,359
? PRIOR FILING DATE: 2000-09-27
? PRIOR APPLICATION NUMBER: PCT/US01/00666
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00667
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00664
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00669
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00665
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00668
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00663
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00662
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00661
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00670
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: US 60/234,687
? PRIOR FILING DATE: 2000-09-21
? PRIOR APPLICATION NUMBER: US 09/608,408
? PRIOR FILING DATE: 2000-06-30
? PRIOR APPLICATION NUMBER: US 09/774,203
? PRIOR FILING DATE: 2001-01-29
? NUMBER OF SEQ ID NOS: 49117
? SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
? SEQ ID NO 25080
? LENGTH: 273
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: MAP TO AL121767.3
? OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
? OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.7
? OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 6.1
? OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
? OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
? OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.5
? OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.5
? OTHER INFORMATION: SWISSPROT HIT: P29539, EVALU6 6.80e-01
? OTHER INFORMATION: EST_HUMAN HIT: BE156318.1, EVALU6 3.00e-78
? OTHER INFORMATION: NT HIT: AL163210.2, EVALU6 7.00e-75
US-09-864-761-25080

Query Match 14.7%; Score 58.4; DB 10; Length 273;
Best Local Similarity 74.0%; Pred. No. 9.4e-09;
Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 297 GGTGAATATTGGCTAATGGAAGCAAAATTTGGCTAATAGAGAAGAAATTTCTAAGCA 356
DB 115 GGGGACATTTGACTGAGAGAGATGATTTGGGATATCTGCTGAGAGAAATTTCTAAGCA 174
QY 357 GCAAAGCATTCAGAGAGTGACTGGCTGCTGTTAAAGGCA 396
DB 175 GCAAAGCATTCAGAGAGTGACTGGCTGCTGTTAAAGGCA 214

RESULT 14
US-09-864-761-8349
; Sequence 8349, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
```

```

? APPLICANT: Hanzel, David K.
? APPLICANT: Chen, Wensheng
? TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
? TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
? FILE REFERENCE: Aeomica-X-1
? CURRENT APPLICATION NUMBER: US/09/864,761
? PRIOR FILING DATE: 2001-05-23
? PRIOR APPLICATION NUMBER: US 60/180,312
? PRIOR FILING DATE: 2000-02-04
? PRIOR APPLICATION NUMBER: US 60/207,456
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: US 09/632,366
? PRIOR FILING DATE: 2000-08-03
? PRIOR APPLICATION NUMBER: GB 24263.6
? PRIOR FILING DATE: 2000-10-04
? PRIOR APPLICATION NUMBER: US 60/236,359
? PRIOR FILING DATE: 2000-09-27
? PRIOR APPLICATION NUMBER: PCT/US01/00666
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00667
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00664
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00669
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00665
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00668
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00663
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00662
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00661
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00670
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: US 60/234,687
? PRIOR FILING DATE: 2000-09-21
? PRIOR APPLICATION NUMBER: US 09/608,408
? PRIOR FILING DATE: 2000-06-30
? PRIOR APPLICATION NUMBER: US 09/774,203
? PRIOR FILING DATE: 2001-01-29
? NUMBER OF SEQ ID NOS: 49117
? SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
? SEQ ID NO 8349
? LENGTH: 519
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: MAP TO AL121767.3
? OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
? OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.7
? OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 6.1
? OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
? OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
? OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.5
? OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.5
US-09-864-761-8349

Query Match 14.7%; Score 58.4; DB 10; Length 519;
Best Local Similarity 74.0%; Pred. No. 1.3e-08;
Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 297 GGTGAATATTGGCTAATGGAAGCAAAATTTGGCTAATAGAGAAGAAATTTCTAAGCA 356
DB 254 GGGGACATTTGACTGAGAGAGATGATTTGGGATATCTGCTGAGAGAAATTTCTAAGCA 313
QY 357 GCAAAGCATTCAGAGAGTGACTGGCTGCTGTTAAAGGCA 396
DB 314 GCAAAGCATTCAGAGAGTGACTGGCTGCTGTTAAAGGCA 353
```

```

RESULT 15
US-09-864-761-2925
: Sequence 2925, Application US/09864761
: Patent No. US2002048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: FILE REFERENCE: Aemica-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
: SEQ ID NO 2925
: LENGTH: 1946
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AC006337.2
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 14
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 15
: OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 13
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 12
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 12
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 12
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 10
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 14
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 13
US-09-864-761-2925
Query Match 14.4%; Score 57; DB 10; Length 1946;
Best Local Similarity 74.2%; Pred. No. 7.3e-08;
Matches 72; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

```

Qy	300	174	360	234
GAATATTTTGGCTATGAGGAACAAATTTTGGCTAAATAAGAGAAATTTTCTAAGCAGCA	GAATTTTGAAGTCTTAAAGAGATGATTTTAGGCTATCTGCTGGAGAAATTTTCTAAGCAGCA	AAGCATTCACAGAGGTGACTGGTGGCTGCTTTAAAGCA	AAGCATTCACAGAGGTGACTGGGTAAGTGTAAAGCA	
359	233	336	270	

Search completed: February 19, 2003, 23:02:47  
Job time : 39.8656 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 16:09:30 : Search time 26.1422 Seconds  
(without alignments)  
4645.318 Million cell updates/sec

Title: US-09-997-610-1\_COPY\_53\_448

Perfect score: 396  
Sequence: 1 gctgactcaccagcacacc.....cttggtctgttaagca 396

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/pdata/2/lna/5A.COMB.seq:\*  
2: /cgn2\_6/pdata/2/lna/5B.COMB.seq:\*  
3: /cgn2\_6/pdata/2/lna/6A.COMB.seq:\*  
4: /cgn2\_6/pdata/2/lna/6B.COMB.seq:\*  
5: /cgn2\_6/pdata/2/lna/6C.COMB.seq:\*  
6: /cgn2\_6/pdata/2/lna/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46.6	11.8	1313	2	US-08-463-911-6 Sequence 6, App1
2	46.6	11.8	4517	4	US-09-140-804-9 Sequence 9, App1
3	42.4	10.7	1107	3	US-09-188-930-217 Sequence 217, App
4	40	10.1	1276	2	US-08-463-911-1 Sequence 1, App1
5	37.4	9.4	728	4	US-09-336-536-2 Sequence 2, App1
6	37.4	9.4	1338	4	US-09-336-536-1 Sequence 1, App1
7	37.4	9.4	1347	4	US-09-140-804-1 Sequence 1, App1
8	36.2	9.1	729	4	US-09-140-804-10 Sequence 10, App1
9	34.4	8.7	14855	2	US-08-687-080-59 Sequence 59, App1
10	34.4	8.7	44453	4	US-09-146-053-5 Sequence 5, App1
11	34.4	8.7	49136	4	US-09-422-869-1 Sequence 1, App1
12	34	8.6	459	4	US-09-413-551-2 Sequence 2, App1
13	33.2	8.4	486	4	US-09-413-551-1 Sequence 1, App1
14	33.2	8.4	1881	4	US-09-426-290-1 Sequence 20, App1
15	33.2	8.4	168575	4	US-09-426-290-1 Sequence 20, App1
16	31.4	7.9	1001	3	US-09-188-930-218 Sequence 218, App
17	31.4	7.9	1015	3	US-09-188-930-30 Sequence 30, App1
18	31.4	7.9	35081	2	US-08-752-760A-1 Sequence 8, App1
19	31	7.8	1215	2	US-09-092-770-8 Sequence 8, App1
20	30.6	7.7	1215	2	US-09-222-851-8 Sequence 8, App1
21	30.6	7.7	2712	3	US-08-346-455B-37 Sequence 37, App1
22	30.6	7.7	2712	3	US-08-977-221-37 Sequence 37, App1
23	30.6	7.7	2712	4	US-09-483-831B-37 Sequence 37, App1
24	30.6	7.7	2712	5	PCT-US95-06613-37 Sequence 37, App1
25	30.6	7.7	2946	1	US-08-346-455B-35 Sequence 35, App1
26	30.6	7.7	2946	3	US-08-977-221-35 Sequence 35, App1
27	30.6	7.7	2946	4	US-09-483-831B-35 Sequence 35, App1

C 28	30.6	7.7	2946	5	PCT-US95-06613-35	Sequence 35, App1
C 29	30.6	7.7	3251	1	US-08-346-455B-68	Sequence 68, App1
C 30	30.6	7.7	3251	3	US-08-977-221-68	Sequence 68, App1
C 31	30.6	7.7	3251	4	US-09-483-831B-68	Sequence 68, App1
C 32	30.6	7.7	3251	5	PCT-US95-06613-68	Sequence 68, App1
C 33	30.4	7.7	420	2	US-08-557-892-5	Sequence 5, App1
C 34	30.4	7.7	420	2	US-08-387-858A-5	Sequence 5, App1
C 35	30.4	7.7	420	4	US-09-294-384B-5	Sequence 5, App1
C 36	30.4	7.7	420	4	US-08-717-079-5	Sequence 5, App1
C 37	30.4	7.7	519	1	US-08-318-905-18	Sequence 18, App1
C 38	30.4	7.7	519	1	US-08-483-232-18	Sequence 18, App1
C 39	30.4	7.7	519	1	US-08-483-140-18	Sequence 18, App1
C 40	30.4	7.7	519	2	US-08-485-938A-18	Sequence 18, App1
C 41	30.4	7.7	519	2	US-08-910-041-18	Sequence 18, App1
C 42	30.4	7.7	519	3	US-09-328-474-18	Sequence 18, App1
C 43	30.4	7.7	519	3	US-09-100-546-18	Sequence 18, App1
C 44	30.4	7.7	519	3	US-09-100-715-18	Sequence 18, App1
C 45	30.4	7.7	519	4	US-09-577-758-18	Sequence 18, App1

## ALIGNMENTS

RESULT 1  
US-08-463-911-6  
; Sequence 6, Application US/08463911  
; Patent No. 5869330  
; GENERAL INFORMATION:  
; APPLICANT: Scherer, Philipp E.  
; APPLICANT: Lodish, Harvey F.  
; TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED  
; TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Millitia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/463,911  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia  
; REGISTRATION NUMBER: 32,227  
; REFERENCE/DOCKET NUMBER: WH195-05  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 861-6240  
; TELEFAX: (617) 861-9540  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1313 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 73..804  
; US-08-463-911-6  
Query Match 11.8%; Score 46.6; DB 2; Length 1313;  
Best Local Similarity 48.9%; Pred. No. 1.8e-05;  
Matches 158; Conservative 0; Mismatches 159; Indels 6; Gaps 1;  
OY 3 TGGACCTCCAGCACACCCAGCCGCCAGAAAGGTGGGGCTCTCTGTCACACGCTTT 62

```
Db 294 TGGTCTTAAGGAGACATCGGTGAACCGGAGTACCGGGGGCTCAAGTCCCGCAGGCTT 353
QY 63 ACCAATATATACAGAGAAATTAAGTGAATGACAAATGCCCCCTGCTCTGATATAGAAG 122
Db 354 TCCGGAAATCCAAAGCAGAGAAAGAAACCTGGAGAAAGTGCCCT-----ATGATACCG 407
QY 123 GTCAGCCTTACTGTGAAGCTCAGTGGAAAACTTCTCTCTTCAAGCCATCATCTT 182
Db 408 CTAGCATTCATGTCGTGGATTTGAGACTTACGTACTATCCCAACATGCCCCATTCGCTT 467
QY 183 CACAGGGGCTCCTGACATGCGCCAGAGGATTTAAAGAGGCGCATGGAGTCTTGCTG 242
Db 468 TACCAAGATCTTCTACATACAGAAACCATATGATGCTCCACTGTAATTCACCTG 527
QY 243 CAGGGTCCCTGGAAATTAATCTCCAGCTTTGATGTTGAGCTGCATTTGCAAGGTGAA 302
Db 528 CAACATTCCTGGGCTGACTACTTGGCTTACCATCATCAGTCTATATGAAGATGTGAA 587
QY 303 TATTTGGCTAATGAGAGCAAA 325
Db 588 GGTACGCTCTCTCAAGAGGACA 610
```

```
RESULT 2
US-09-140-804-9
; Sequence 9, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: ADIPOCTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49
; CURRENT APPLICATION NUMBER: US/09/140,804
; EARLIER FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: 60/056,983
; EARLIER FILING DATE: 1997-08-26
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 4517
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-140-804-9
```

```
Query Match 11.8%; Score 46.6; DB 4; Length 4517;
Best Local Similarity 48.9%; Pred. No. 3.3e-05;
Matches 158; Conservative 0; Mismatches 159; Indels 6; Gaps 1;
```

```
QY 3 TGGACTCCAGACACCCCAAGGAGAAAGTGGGCTCTCTGTCACCAAGCTTT 62
Db 248 TGGTCCCTTAAGGAGACATCGGTGAACCGGAGTACCGGGCTCAAGTCCCGAGGCTT 307
QY 63 ACCAATATATACAGAGAAATTAAGTGAATGACAAATGCCCCCTGCTCTGATATAGAAG 122
Db 308 TCCGGAAATCCAAAGCAGAGAAAGAAACCTGGAGAAAGTGCCCT-----ATGATACCG 361
QY 123 GTCAGCCTTACTGTGAAGCTCAGTGGAAAACTTCTCTCTTCAAGCCATCATCTT 182
Db 362 CTCAGCATTCAGTGTGGATTTGAGACTTACGTACTATCCCAACATGCCCCATTCGCTT 421
QY 183 CACAGGGGCTCCTGACATGCGCCAGAGGATTTAAAGAGGCGCATGGAGTCTTGCTTG 242
Db 422 TACCAAGATCTTCTACATACAGAAACCATATGATGCTCCACTGTAATTCACCTG 481
QY 243 CAGGGTCCCTGGAAATTAATCTCCAGCTTTGATGTTGAGCTGCATTTGCAAGGTGAA 302
Db 481 CAACATTCCTGGGCTGACTACTTGGCTTACCATCATCAGTCTATATGAAGATGTGAA 541
QY 303 TATTTGGCTAATGAGAGCAAA 325
Db 542 GGTACGCTCTCTCAAGAGGACA 564
```

```
RESULT 3
US-09-188-930-217
; Sequence 217, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; EARLIER FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 217
; LENGTH: 1107
; TYPE: DNA
; ORGANISM: Rat
US-09-188-930-217
```

```
Query Match 10.7%; Score 42.4; DB 3; Length 1107;
Best Local Similarity 50.5%; Pred. No. 0.00038;
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
```

```
QY 117 AGAAGGTACACCTTTACTGTGAAGCTCAGTGAAGAAACTCTCTCTTCAAGCCCAT 176
Db 614 AGCCAAGTGGCTTTTCTGCGGCGGTACCAAGNATTACCACTGAGCCACTGCCCCAT 673
QY 177 CATCTTCACAGGGTCTCTGTCATATGCCAGAGGATTTAAAGAGGCCATGGAGTCTT 236
Db 674 CAAGTTTGACAAGATTTGATGAAATGAGGAGGCGCACATCAATGCATCAGTGGCAAGTT 733
QY 237 TGTTCAGAGGTGCTGGGAATTAATCTCCAGCTTTGATGTTGAGCTGCATCATTTGCA 296
Db 734 CGTCTGACGCTGCCAGGATCTATTTACTTACCTATGACATTAACGTGGCAACAAACA 793
QY 297 GGTGATATTTGGCTAATGAGGAA 320
Db 794 CCTGGCCATCGGCTTACTGTCACAA 817
```

```
RESULT 4
US-08-463-911-1
; Sequence 1, Application US/08463911
; Patent No. 5869330
; GENERAL INFORMATION:
; APPLICANT: Scherer, Philipp E.
; APPLICANT: Lodish, Harvey F.
; TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
; TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCTES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,911
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
```







```
RESULT 15
US-09-426-290-1/C
; Sequence 1, Application us/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Berglund Ran Olafsdottir
; APPLICANT: Jeffrey Gulcher
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/09/426,290
; CURRENT FILING DATE: 1999-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 168575
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21181)...(21403)
; NAME/KEY: CDS
; LOCATION: (95252)...(95430)
; NAME/KEY: CDS
; LOCATION: (101753)...(101996)
; NAME/KEY: CDS
; LOCATION: (110324)...(110439)
; NAME/KEY: CDS
; LOCATION: (124058)...(124278)
; NAME/KEY: CDS
; LOCATION: (127009)...(127130)
; NAME/KEY: CDS
; LOCATION: (128910)...(129139)
US-09-426-290-1

Query Match      8.4%; Score 33.2; DB 4; Length 168575;
Best Local Similarity 75.9%; Pred. No. 4.9;
Matches 41; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 336 TAAGGAGAATTTCTAAGCAGCAAGCATTCAAGAGCTTGGGTGCTGTT 389
    || ||||| |||| ||||| ||||| || ||||| ||||| ||||| |||||
Db 64084 TATGAGAGAATTTTAAAGAGCAAGCATTCAATGATGCTGTTGGCTACTTTT 64031

Search completed: February 19, 2003, 23:02:25
Job time : 231.142 secs
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 15:08:07 ; Search time 137.053 Seconds  
(without alignments)  
6506.903 Million cell updates/sec

Title: US-09-997-610-1\_COPY\_53\_448

Perfect score: 396  
Sequence: 1 gctgagctccagcacacc.....cttggtgtgttaagca 396

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :  
1: N.Geneseq\_101002.\*  
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
25: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	351	88.6	1338	24	AA144066
2	58.6	14.8	404	21	AAC24351
3	58.4	14.7	273	22	ABA75076
4	58.4	14.7	273	22	ABA39760
5	58.4	14.7	273	22	AAK23591
6	58.4	14.7	273	22	AAK49719
7	58.4	14.7	273	22	AA126826
8	58.4	14.7	273	22	AA155610
9	58.4	14.7	519	22	ABA62555

10	58.4	14.7	519	22	ABA29883	Probe #8349 for ge
11	58.4	14.7	519	22	AAK10911	Human brain expres
12	58.4	14.7	519	22	AAK36769	Human bone marrow
13	58.4	14.7	519	22	AA117621	Probe #7554 for ge
14	58.4	14.7	519	22	AA142541	Probe #11227 used
15	57	14.4	662	22	AA115063	Human breast cancer
16	57	14.4	686	22	AA123916	Human breast cancer
17	57	14.4	920	23	AA181134	DNA encoding novel
18	57	14.4	999	23	AA566807	DNA encoding novel
19	57	14.4	1077	23	AA582659	DNA encoding novel
20	57	14.4	1946	22	ABA44227	Human breast cell
21	57	14.4	1946	22	ABA54677	Human foetal liver
22	57	14.4	1946	22	ABA24459	Probe #2925 for ge
23	57	14.4	1946	22	AAK02964	Human brain expres
24	57	14.4	1946	22	AAK28410	Human bone marrow
25	57	14.4	1946	22	AA112975	Probe #2908 for ge
26	57	14.4	1946	22	AA134334	Probe #3020 used t
27	57	14.4	1946	22	AA102893	Probe #2884 used t
28	57	14.4	1946	24	AB502918	Human genome-deriv
29	56.4	14.2	32127	22	AA199255	Human excretory re
30	56.4	14.2	32127	22	AA163605	Human kidney relat
31	56.4	14.2	72215	22	AAK86832	Human immune/haema
32	56	14.1	505	22	ABA60370	Human foetal liver
33	56	14.1	505	22	ABA28609	Probe #7075 for ge
34	56	14.1	505	22	AAK08648	Human brain expres
35	56	14.1	505	22	AAK34535	Human bone marrow
36	56	14.1	505	22	AA140252	Probe #8938 used t
37	56	14.1	505	24	AB509176	Human genome-deriv
38	56	14.1	1054	23	AA564468	DNA encoding novel
39	55.4	14.0	342	21	AA266950	Human secreted pro
40	55.4	14.0	520	22	ABA62906	Human foetal liver
41	55.4	14.0	520	22	ABA30177	Probe #8643 for ge
42	55.4	14.0	520	22	AAK11311	Human brain expres
43	55.4	14.0	520	22	AAK37101	Human bone marrow
44	55.4	14.0	520	22	AA117942	Probe #7875 for ge
45	55.4	14.0	520	22	AA142928	Probe #11614 used

## ALIGNMENTS

RESULT 1  
ID AAL44066  
AAL44066 standard; cDNA; 1338 BP.  
AC AAL44066;  
XX  
DT 27-SEP-2002 (first entry)  
XX  
DE Human genset metabolic gene (GMC-9) cDNA sequence.  
XX  
KW Human; gene; ss; gene therapy; genset metabolic gene; GMC-7A; GMC-7B;  
KW GMC-8; GMC-9; GMC-10; GMC-11; metabolic-related disorder; obesity;  
KW Impaired glucose tolerance; insulin resistance; Syndrome X;  
KW Type II diabetes; hyperlipidaemia; atherosclerosis; hypertension;  
KW heart disease; cardiac insufficiency; coronary insufficiency;  
KW high blood pressure; insulin sensitizer;  
KW non-insulin dependent diabetes mellitus.  
XX  
OS Homo sapiens.  
XX  
FH Key 1.1338 Location/Qualifiers  
FT CDS  
FT  
FT /tag= a  
FT /partial  
FT /product= "Human GMC-9 protein"  
FT /note= "No stop codon is given"  
PN WO200255694-A2.  
XX  
PD 18-JUL-2002.  
XX  
PF 15-JAN-2002; 2002WO-IB01215.



PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483447/52.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human fetal liver -  
XX  
XX  
XX Claim 4; SEQ ID NO 23381; 639pp + sequence listing; English.  
XX  
XX The invention relates to a single exon nucleic acid probe for  
XX measuring human gene expression in a sample derived from human foetal  
XX liver. The single exon nucleic acid probes may be used for predicting,  
XX measuring and displaying gene expression in samples derived from human  
XX fetal liver. The present sequence is a single exon nucleic acid  
XX probe of the invention.  
XX Note: The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ffp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence 273 BP; 80 A; 36 C; 81 G; 76 T; 0 other;  
  
Query Match 14.7%; Score 58.4; DB 22; Length 273;  
Best Local Similarity 74.0%; Pred. No. 7.6e-09;  
Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
  
QY 297 GGTGAATATTGGCTAATGAGCAAAATTTGGCTAATAGAGAAATTTCTAAGCA 356  
DB 115 GGGGAACTTGAACTTGAGAGATGATTTGGGGTATCTGTGGAGAAATTTCTAACA 174  
  
QY 357 GCAAGCATTCAGAGGTGACTGGGCTGCTTAAGGCA 396  
DB 175 GCAAGCATTCAGAGGTGACTGGCTGCTTAAGGCA 214  
  
RESULT 4  
ABA39760  
ID ABA39760 standard; DNA; 273 BP.  
XX  
XX ABA39760;  
XX  
XX  
XX 23-JAN-2002 (first entry)  
XX  
XX Probe #18226 for gene expression analysis in human heart cell sample.  
XX  
XX Human; gene expression; heart; microarray; vascular system; probe;  
XX cardiovascular disease; hypertension; cardiac arrhythmia;  
XX congenital heart disease; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200157274-A2.  
XX  
XX  
XX 09-AUG-2001.  
XX  
XX  
XX 30-JAN-2001; 2001WO-US00666.  
XX  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.

XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-48899/53.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX hearts -  
XX  
XX  
XX Claim 4; SEQ ID NO 18226; 530pp; English.  
XX  
XX The present invention relates to single exon nucleic acid probes for  
XX measuring human gene expression in a sample derived from human heart. The  
XX present sequence is one such probe. The probes may be used for  
XX predicting, measuring and displaying gene expression in samples derived  
XX from the human heart via microarrays. By measuring gene expression, the  
XX probes are useful for predicting, diagnosing, grading, staging,  
XX monitoring and prognosing diseases of the human heart and vascular system  
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
XX congenital heart disease.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ffp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence 273 BP; 80 A; 36 C; 81 G; 76 T; 0 other;  
  
Query Match 14.7%; Score 58.4; DB 22; Length 273;  
Best Local Similarity 74.0%; Pred. No. 7.6e-09;  
Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
  
QY 297 GGTGAATATTGGCTAATGAGCAAAATTTGGCTAATAGAGAAATTTCTAAGCA 356  
DB 115 GGGGAACTTGAACTTGAGAGATGATTTGGGGTATCTGTGGAGAAATTTCTAACA 174  
  
QY 357 GCAAGCATTCAGAGGTGACTGGGCTGCTTAAGGCA 396  
DB 175 GCAAGCATTCAGAGGTGACTGGCTGCTTAAGGCA 214  
  
RESULT 5  
AAK23591  
ID AAK23591 standard; DNA; 273 BP.  
XX  
XX AAK23591;  
XX  
XX  
XX 05-NOV-2001 (first entry)  
XX  
XX Human brain expressed single exon probe SEQ ID NO: 23582.  
XX  
XX Human; brain expressed exon; gene expression analysis; probe;  
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
XX epilepsy; cancer; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200157275-A2.  
XX  
XX  
XX 09-AUG-2001.  
XX  
XX  
XX 30-JAN-2001; 2001WO-US00667.  
XX  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483446/52.  
XX

XX Single exon nucleic acid probes for analyzing gene expression in human  
 PT brains -  
 XX  
 PS Example 4; SEQ ID NO: 23582; 650pp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC brain. They can be used to measure gene expression in brain cell samples,  
 CC which may enable the diagnosis and improved treatment of nervous system  
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 CC epilepsy and cancers. The present sequence is one of the probes of the  
 CC invention.  
 CC  
 SQ Sequence 273 BP; 80 A; 36 C; 81 G; 76 T; 0 other;  
 XX  
 Query Match 14.7%; Score 58.4; DB 22; Length 273;  
 Best Local Similarity 74.0%; Pred. No. 7.6e-09;  
 Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
 XX  
 QY 297 GGTGAATATTTGGCTAATGAGCAAAATTTGGCTAATAGAGAAATTTCTAAGCA 356  
 DB 115 GGGGAACATTGACTTGAGAGAGATGATTTGGGTATCTGGTGAAGAAATTTCTAACA 174  
 XX  
 QY 357 GCAAGCATTCAAGAGGTGACTTGGCTGCTTTAAGGCA 396  
 DB 175 GCAAGCATTCAAGAGGTGACTTGGCTGCTTTAAGGCA 214  
 XX  
 RESULT 6  
 ID AAK49719 standard; DNA: 273 BP.  
 XX  
 AC AAK49719;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human bone marrow expressed single exon probe SEQ ID NO: 24276.  
 XX  
 KW Human; bone marrow expressed exon; gene expression analysis; probe;  
 KM microarray; cancer; Leukemia; Lymphoma; myeloma; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157276-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00668.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 XX  
 PR 26-MAY-2000; 2000US-0207456.  
 XX  
 PR 30-JUN-2000; 2000US-0608408.  
 XX  
 PR 03-AUG-2000; 2000US-0632366.  
 XX  
 PR 21-SEP-2000; 2000US-0234687.  
 XX  
 PR 27-SEP-2000; 2000US-0236359.  
 XX  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488900/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human bone marrow -  
 XX  
 PS Example 4; SEQ ID NO: 24276; 658pp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers

CC such as lymphoma, leukemia and myeloma. The present sequence is one of  
 CC the probes of the invention.  
 XX  
 SQ Sequence 273 BP; 80 A; 36 C; 81 G; 76 T; 0 other;  
 XX  
 Query Match 14.7%; Score 58.4; DB 22; Length 273;  
 Best Local Similarity 74.0%; Pred. No. 7.6e-09;  
 Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
 XX  
 QY 297 GGTGAATATTTGGCTAATGAGCAAAATTTGGCTAATAGAGAAATTTCTAAGCA 356  
 DB 115 GGGGAACATTGACTTGAGAGAGATGATTTGGGTATCTGGTGAAGAAATTTCTAACA 174  
 XX  
 QY 357 GCAAGCATTCAAGAGGTGACTTGGCTGCTTTAAGGCA 396  
 DB 175 GCAAGCATTCAAGAGGTGACTTGGCTGCTTTAAGGCA 214  
 XX  
 RESULT 7  
 ID AAI26826 standard; DNA: 273 BP.  
 XX  
 AC AAI26826;  
 XX  
 DT 12-OCT-2001 (first entry)  
 XX  
 DE Probe #16759 for gene expression analysis in human cervical cell sample.  
 XX  
 KW Probe; human; microarray; gene expression; cervical epithelial cell;  
 KM cervical cancer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157278-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00670.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 XX  
 PR 26-MAY-2000; 2000US-0207456.  
 XX  
 PR 30-JUN-2000; 2000US-0608408.  
 XX  
 PR 03-AUG-2000; 2000US-0632366.  
 XX  
 PR 21-SEP-2000; 2000US-0234687.  
 XX  
 PR 27-SEP-2000; 2000US-0236359.  
 XX  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488901/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human cervical epithelial cells -  
 XX  
 PS Claim 25; SEQ ID NO 16759; 487pp; English.  
 XX  
 CC The present invention relates to human single exon nucleic acid probes  
 CC (SENPs). The present sequence is one such probe. The SENPs are derived  
 CC from human HeLa cells. The SENPs can be used to produce a single exon  
 CC microarray, which can be used for measuring human gene expression in a  
 CC sample derived from human cervical epithelial cells. By measuring gene  
 CC expression, the probes are therefore useful in grading and/or staging  
 CC of diseases of the cervix, notably cervical cancer.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 XX  
 SQ Sequence 273 BP; 80 A; 36 C; 81 G; 76 T; 0 other;  
 XX  
 Query Match 14.7%; Score 58.4; DB 22; Length 273;  
 Best Local Similarity 74.0%; Pred. No. 7.6e-09;

[illegible]

KW Human; gene expression; heart; microarray; vascular system; probe;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease; ss.  
OS Homo sapiens.  
XX WO200157274-A2.  
PN 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US00666.  
PD 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-488899/53.  
XX Single exon nucleic acid probes for analyzing gene expression in human hearts -  
PT  
PT  
XX  
XX  
PS Claim 1; SEQ ID No 8349; 530pp; English.  
CC The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart. The  
CC present sequence is one such probe. The probes may be used for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from the human heart via microarrays. By measuring gene expression, the  
CC probes are useful for predicting, diagnosing, grading, staging,  
CC monitoring and prognosing diseases of the human heart and vascular system  
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
CC congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 519 BP; 170 A; 68 C; 148 G; 133 T; 0 other;  
Query Match 14.7%; Score 58.4; DB 22; Length 519;  
Best Local Similarity 74.0%; Pred. No. 1.1e-08;  
Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
QY 297 GGTGAATATTGGCTAATGAGGAGCAAAATTTGGCTAATAAGGAGAAATTTCTTAAGCA 356  
DB 254 GGGGAACATTGAACCTGAGAGAGATGATTTGGGGTATCTGCTGGTGAAGAAATTTCTTAACA 313  
QY 357 GCAAGCATTCAAGAGGTGACTGGTGGTCTGCTTTAAAGGCA 396  
DB 314 GCAAGCATTCAAGAGGTGACTGGTGGTCTGCTTTAAAGGCA 353  
RESULT 11  
AAK10911  
ID AAK10911 standard; DNA; 519 BP.  
XX  
AC AAK10911;  
XX  
XX 05-NOV-2001 (first entry)  
DE Human brain expressed single exon probe SEQ ID NO: 10902.  
XX  
XX Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer; ss.  
XX

OS Homo sapiens.  
XX WO200157275-A2.  
XX  
PD 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US00667.  
XX 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-483446/52.  
XX Single exon nucleic acid probes for analyzing gene expression in human brains -  
DR  
DR  
PT  
PT  
XX  
XX  
PS Example 4; SEQ ID NO: 10902; 650pp + Sequence Listing; English.  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention.  
XX  
SQ Sequence 519 BP; 170 A; 68 C; 148 G; 133 T; 0 other;  
Query Match 14.7%; Score 58.4; DB 22; Length 519;  
Best Local Similarity 74.0%; Pred. No. 1.1e-08;  
Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
QY 297 GGTGAATATTGGCTAATGAGGAGCAAAATTTGGCTAATAAGGAGAAATTTCTTAAGCA 356  
DB 254 GGGGAACATTGAACCTGAGAGAGATGATTTGGGGTATCTGCTGGTGAAGAAATTTCTTAACA 313  
QY 357 GCAAGCATTCAAGAGGTGACTGGTGGTCTGCTTTAAAGGCA 396  
DB 314 GCAAGCATTCAAGAGGTGACTGGTGGTCTGCTTTAAAGGCA 353  
RESULT 12  
AAK36769  
ID AAK36769 standard; DNA; 519 BP.  
XX  
AC AAK36769;  
XX  
XX 06-NOV-2001 (first entry)  
DT  
XX  
DE Human bone marrow expressed single exon probe SEQ ID NO: 11326.  
XX  
XX Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
XX  
XX Homo sapiens.  
XX  
PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00668.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:16:26 ; Search time 1262.56 Seconds  
(without alignments)  
9128.055 Million cell updates/sec

Title: US-09-997-610-1-copy\_53\_448

Perfect score: 396

Sequence: 1 gctggacctccagcacacc.....cttgggtgctgttaaggca 396

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_on.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pi.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	351	88.6	145880	9	HS302D9	282198 Human DNA s
2	146.2	36.9	1265	10	TMSHP20A	DI2974 Tamlas asia
3	139.2	35.2	5121	10	AB067779	AB067779 Tamlas si
4	83	21.0	2005	10	TMSHP25	DI2975 Tamlas asia
5	68.6	17.3	559	10	AB067813s3	AB069679 Callosclu
6	68.2	17.2	145880	9	HS302D9	282198 Human DNA s
7	63.4	16.0	90497	9	HS73284	AL008722 Human DNA
8	63.4	16.0	135880	2	AC104256	AC104256 Homo sapi
9	63.2	16.0	51000	9	AP005265	AP005265 Homo sapi
10	63.2	16.0	144514	2	AC027433	AC027433 Homo sapi
11	63.2	16.0	149038	2	AC025112	AC025112 Homo sapi
12	62.6	15.8	162063	9	AC015553	AC015553 Homo sapi
13	62.6	15.8	175134	2	AL161623	AL161623 Homo sapi
14	62.6	15.8	176932	9	AL157884	AL157884 Human DNA
15	62.4	15.8	1385	10	TMSHP27	DI2976 Tamlas asia
16	62	15.7	203234	9	AL441943	AL441943 Human DNA
17	61.8	15.6	69937	2	AC069566	AC069566 Homo sapi
18	61.8	15.6	122231	2	AC026099	AC026099 Homo sapi
19	61.8	15.6	143872	9	AC091987	AC091987 Homo sapi
20	61.8	15.6	172792	9	AC062031	AC062031 Homo sapi
21	61.8	15.6	179848	9	AC092967	AC092967 Homo sapi
22	61.8	15.6	180038	2	AC026870	AC026870 Homo sapi
23	61.2	15.5	177386	2	AC018971	AC018971 Homo sapi
24	60.2	15.2	108511	9	AC011418	AC011418 Homo sapi
25	60.2	15.2	137591	2	AL672086	AL672086 Homo sapi
26	60.2	15.2	158587	9	AC074325	AC074325 Homo sapi
27	60.2	15.2	158811	2	AC013503	AC013503 Homo sapi
28	60.2	15.2	160151	9	AC098595	AC098595 Homo sapi
29	60.2	15.2	161049	2	AL591842	AL591842 Homo sapi
30	60.2	15.2	173767	9	AC002449	AC002449 Human PAC
31	60.2	15.2	175765	9	AC027128	AC027128 Homo sapi
32	60.2	15.2	179484	9	AC107626	AC107626 Homo sapi
33	60.2	15.2	187201	2	AC068063	AC068063 Homo sapi
34	60.2	15.2	187640	9	AC073840	AC073840 Homo sapi
35	60.2	15.2	188622	9	AL359258	AL359258 Human DNA
36	60.2	15.2	192973	9	AC023471	AC023471 Homo sapi
37	60.2	15.2	207600	9	AC024910	AC024910 Homo sapi
38	60.2	15.2	222542	9	AC022379	AC022379 Homo sapi
39	59.8	15.1	170200	9	AC008012	AC008012 Homo sapi
40	59.6	15.1	144057	2	AC074266	AC074266 Homo sapi
41	59.6	15.1	172349	9	AC019185	AC019185 Homo sapi
42	59.6	15.1	195594	2	AC096992	AC096992 Homo sapi
43	59.6	15.1	197078	2	AC117382	AC117382 Homo sapi
44	59.6	15.1	233877	9	AC093798	AC093798 Homo sapi
45	59.4	15.0	98070	9	AC025463	AC025463 Homo sapi

ALIGNMENTS

RESULT 1  
HS302D9  
LOCUS  
DEFINITION Human DNA sequence from clone RP1-302D9 on chromosome 22 Contains  
GSSs, complete sequence.  
ACCESSION 282198  
VERSION 282198.2 GI:6572207  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 145880)  
AUTHORS Bridgeman,A.  
TITLE Direct Submission

## JOURNAL

Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Dec 13, 1999 this sequence version replaced gi:3164067.  
During sequence assembly this data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.  
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at  
<http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr22>  
RP1-302D9 is from the library RPci-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/> VECTOR: PCPAC2

This sequence is the entire insert of clone RP1-302D9 The true left end of clone CTA-282F2 is at 69682 in this sequence. The true right end of clone CTA-415G2 is at 55167 in this sequence.

## FEATURES

Source	Location/Qualifiers
	1..145880 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="22" /clone="RP1-302D9" /clone_lib="RPci-1"
repeat_region	188..245 /note="MER3 repeat: matches 144..209 of consensus"
repeat_region	246..571 /note="AluX repeat: matches 1..312 of consensus"
repeat_region	572..759 /note="MER3 repeat: matches 1..144 of consensus"
repeat_region	783..933 /note="MER5A repeat: matches 26..187 of consensus"
repeat_region	1033..1336 /note="AluSp repeat: matches 1..299 of consensus"
repeat_region	1450..1583 /note="MIR repeat: matches 24..160 of consensus"
repeat_region	1687..1752 /note="L2 repeat: matches 2593..2661 of consensus"
repeat_region	2350..2660 /note="AluSc repeat: matches 3..309 of consensus"
repeat_region	2684..2981 /note="AluSq repeat: matches 2..300 of consensus"
repeat_region	3323..3343 /note="MLT1E repeat: matches 116..136 of consensus"
repeat_region	3344..3652 /note="AluY repeat: matches 1..309 of consensus"
repeat_region	3653..3928 /note="MLT1E repeat: matches 136..359 of consensus"
repeat_region	3929..4278 /note="RHE1B repeat: matches 3..364 of consensus"
repeat_region	4279..4485 /note="MLT1E repeat: matches 359..568 of consensus"
repeat_region	5073..5176 /note="52 copies 2 mer ct 78 conserved"
repeat_region	5181..5491 /note="AluDb repeat: matches 1..311 of consensus"
repeat_region	6369..6485 /note="L2 repeat: matches 2579..2705 of consensus"
repeat_region	6647..6685 /note="MADE1 repeat: matches 1..23 of consensus"
repeat_region	6686..6987 /note="AluX repeat: matches 1..302 of consensus"
repeat_region	6988..7036 /note="MADE1 repeat: matches 23..77 of consensus"
repeat_region	7482..7754 /note="AluJb repeat: matches 9..290 of consensus"
repeat_region	7775..8060 /note="AluJo repeat: matches 1..295 of consensus"
repeat_region	8414..8551 /note="L2 repeat: matches 2553..2706 of consensus"
repeat_region	8914..9030 /note="MIR repeat: matches 147..262 of consensus"
repeat_region	9110..9280 /note="MIR repeat: matches 91..262 of consensus"
repeat_region	9283..9412 /note="MIR repeat: matches 15..144 of consensus"
repeat_region	9521..9679 /note="FAM repeat: matches 3..161 of consensus"
repeat_region	9820..10225 /note="MSTB repeat: matches 2..425 of consensus"
misc_feature	complement(10179..10678) /note="match: GSS: Em:B56592"
misc_feature	complement(10204..10728) /note="match: GSS: Em:AQ701486"
misc_feature	complement(10249..10706) /note="match: GSS: Em:AQ225495"
repeat_region	10312..10383 /note="MIR repeat: matches 79..150 of consensus"
misc_feature	10718..11310 /note="match: GSS: Em:B14024"
misc_feature	10784..11201 /note="match: GSS: Em:B43656"
repeat_region	11838..11946 /note="MIR repeat: matches 20..137 of consensus"
repeat_region	12174..12445 /note="L2 repeat: matches 1988..2275 of consensus"
repeat_region	12444..12642 /note="MIR repeat: matches 63..241 of consensus"
misc_feature	13017..13369 /note="match: STS: Em:G49301"
repeat_region	13331..13397 /note="MIR repeat: matches 174..244 of consensus"
repeat_region	13398..13698 /note="AluSp repeat: matches 1..302 of consensus"
repeat_region	13699..13810 /note="MIR repeat: matches 76..174 of consensus"
repeat_region	13806..13919 /note="MIR repeat: matches 77..189 of consensus"
repeat_region	13945..14060 /note="MIR repeat: matches 24..142 of consensus"
repeat_region	14061..14367 /note="AluY repeat: matches 1..301 of consensus"
repeat_region	14368..14452 /note="MIR repeat: matches 141..225 of consensus"
repeat_region	14589..14679 /note="MIR repeat: matches 173..262 of consensus"
misc_feature	14597..15201 /note="match: GSS: Em:AQ553482"
misc_feature	14616..15060 /note="match: GSS: Em:AQ370601"
repeat_region	14868..15040 /note="MIR repeat: matches 49..233 of consensus"
repeat_region	15071..15188 /note="L2 repeat: matches 2112..2239 of consensus"
repeat_region	15304..15399 /note="MLT1B repeat: matches 1..99 of consensus"
repeat_region	15490..15662 /note="AluSg1 repeat: matches 2..114 of consensus"
repeat_region	15669..15727 /note="MLT1B repeat: matches 119..178 of consensus"
repeat_region	15728..16027

repeat\_region /note="AluSc repeat: matches 1. .299 of consensus" 16028. .16245  
repeat\_region /note="MLT1B repeat: matches 178. .390 of consensus" 16546. .16854  
repeat\_region /note="AluY repeat: matches 1. .300 of consensus" 18296. .18323  
repeat\_region /note="MSTA repeat: matches 2. .29 of consensus" 18324. .18392  
repeat\_region /note="MER66-internal repeat: matches 4919. .4993 of consensus" 18393. .18712  
repeat\_region /note="AluJb repeat: matches 1. .311 of consensus" 18713. .19133  
repeat\_region /note="MER66-internal repeat: matches 4548. .4919 of consensus" complement(18872. .19230)  
misc\_feature /note="match: GSS: Em:AQ005063" 19251. .19719  
misc\_feature /note="match: GSS: Em:B14179" 19537. .20290  
repeat\_region /note="HERVH21 repeat: matches 4657. .5784 of consensus" 20317. .20382  
repeat\_region /note="33 copies 2 mer ta 68 conserved" 20513. .20666  
repeat\_region /note="77 copies 2 mer tt 70 conserved" 20682. .21008  
repeat\_region /note="AluSg1 repeat: matches 1. .306 of consensus" 21239. .21553  
repeat\_region /note="HUEB-P3 repeat: matches 4410. .4713 of consensus" 21882. .22254  
repeat\_region /note="THE1B repeat: matches 1. .364 of consensus" 22302. .22537  
repeat\_region /note="MER66-internal repeat: matches 2186. .2417 of consensus" 22538. .22850  
repeat\_region /note="AluSp repeat: matches 1. .313 of consensus" 22851. .23801  
repeat\_region /note="MER66-internal repeat: matches 1210. .2186 of consensus" 23905. .23989  
repeat\_region /note="MER66-internal repeat: matches 3017. .3102 of consensus" 88.6%; Score 351; DB 9; Length 145880;  
Query Match Best Local Similarity 98.6%; Pred. No. 2.4e-94;  
Matches 354; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 38 TGGGGCTCTGCTGATATAGAAAGTCAAGCTTACTGTGAAGCTCAGTGGAAACTTC 157  
DB 37254 TGGATGCTCTGCTGATATAGAAAGTCAAGCTTACTGTGAAGCTCAGTGGAAACTTC 37313  
QY 98 AATGCCCTCTGCTGATATAGAAAGTCAAGCTTACTGTGAAGCTCAGTGGAAACTTC 157  
DB 37314 AATGCCCTCTGCTGATATAGAAAGTCAAGCTTACTGTGAAGCTCAGTGGAAACTTC 37373  
QY 158 CTTCTCTTCAAGCCCATCTTCACAGGGTCTGTACAAATGCCAGAGGGATTAA 217  
DB 37374 CTTCTCTTCAAGCCCATCTTCACAGGGTCTGTACAAATGCCAGAGGGATTAA 37433  
QY 218 AGAGGCCATGGAGTCTTGTGTCAGGGTGCCTGGGAATTACTACTCCAGTTGATG 277  
DB 37434 AGAGGCCATGGAGTCTTGTGTCAGGGTGCCTGGGAATTACTACTCCAGTTGATG 37493  
QY 278 TTGAGCTGCATCTTCAAGAGTGAATATTTGGCTAATGAGGAAGCAAAATTTGGCTAATA 337  
DB 37494 TTGAGCTGCATCTTCAAGAGTGAATATTTGGCTAATGAGGAAGCAAAATTTGGCTAATA 37553  
QY 338 AGAAGAAATTTCTAAGAGCAAGCAATTCACAGAGTGTCTGGGTGCTGTTAAAGGCA 396  
DB 37554 AGAAGAAATTTCTAAGAGCAAGCAATTCACAGAGTGTCTGGGTGCTGTTAAAGGCA 37612

RESULT 2  
TMSHP20A

LOCUS TMSHP20A 1265 bp mRNA linear ROD 03-FEB-1999  
DEFINITION Tamias asiaticus mRNA for HP-20, complete cds.  
ACCESSION D12974  
VERSION D12974.1 GI:287467  
KEYWORDS HP-20; collagen-like domain; hibernation-related protein; plasma protein.  
SOURCE Tamias asiaticus liver cDNA to mRNA, clone: pCM20-7.  
ORGANISM Tamias sibiricus  
Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Scurinae; Tamias.  
REFERENCE 1 (bases 1 to 1265)  
AUTHORS Takamatsu,N., Ohba,K., Kondo,J., Kondo,N. and Shiba,T.  
TITLE Hibernation-associated gene regulation of plasma proteins with a collagen-like domain in mammalian hibernators  
JOURNAL Mol. Cell. Biol. 13 (3), 1516-1521 (1993)  
MEDLINE 93180798  
REFERENCE 2 (bases 1 to 1265)  
AUTHORS Takamatsu,N.  
TITLE Direct Submission  
JOURNAL Submitted (21-AUG-1992) Nobuhiko Takamatsu, School of Hygienic Sciences, Kitasato University, Laboratory of Molecular Biology; 1-15-1 Kitasato, Sagamihara, Kanagawa 228, Japan (Tel:0427-78-9408, Fax:0427-78-9403)  
FEATURES  
Location/Qualifiers  
1. .1265  
/organism="Tamias sibiricus"  
/db\_xref="taxon:64680"  
/clone="pCM20-7"  
/tissue\_type="liver"  
89. .679  
/note="collagen-like domain at nt 161-277"  
/codon\_start=1  
/product="HP-20"  
/protein\_id="BAA02351.1"  
/db\_xref="GI:287468"  
/translation="MTDVMRLAIFVLMVNLNDQVSCSPGPGVGVGVPVGPGRGP  
PQCGAAGRPDGPCKPSVKPCRSAPTVKFSRLPSEPVPVFTLVNTQDRL  
KESTGVNCFEPNGYHFSFDVELYHCKVIGLKNHIOVMEKHQLSKNEYENASGAMI  
MPLRGDKVWLEADVETEEPDQAKVIVFSGFLISS"  
sig\_peptide 89. .157  
mat\_peptide 158. .676  
polyA\_signal /product="HP-20"  
polyA\_site 1238. .1243  
BASE COUNT 354 a 304 c 288 g 319 t  
ORIGIN  
Query Match 36.9%; Score 146.2; DB 10; Length 1265;  
Best Local Similarity 68.0%; Pred. No. 7e-33;  
Matches 221; Conservative 0; Mismatches 98; Indels 6; Gaps 1;  
QY 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTGGTGACACAGGTT 61  
DB 186 CAGGAGTCCCGAGTGGCGGGGCTCGTGGCCCCCAGGTCAACAGGCGCTGCGGGTA 245  
QY 62 TACCAATATACAGGAGAAATGAATGAAATGACAAATGCCCTGTCTGTATATAGAA 121  
DB 245 GGCCAGGAGACCAGGACCAAAAGGCCACAGTGTAAATGCCCTGC-----AGAGA 299  
QY 122 GGTACGCTTTACTGTGAAGCTCAGTGGAAACTTCTCTTCTTTCAAGCCCATCT 181  
DB 300 GGTACGCTTTCAGGTTGAAGTTCAGGGAAGGCTCGCTCCACCTTCGGAGCTGTGTCT 359  
QY 182 TCACAGGGTCTGTACAATGCCAGAGGATTTAAGGAGGCATGCGGAGCTTTGCTT 241  
DB 360 TCACAGGCTCTGTACAATGCCAGGACCTTGAAGGAGAGCAGCAGGAGTCTTTAACT 419  
QY 242 GCAGGCTCTGGAAATTAATCTACTCCAGCTTTGATGCTGAGTGCATCATTTGCAAGTGA 301  
DB 420 GCGTGGAGCTGGAAATTAATCTACTCCAGCTTTGATGCTGAGTGCATCATTTGCAAGTGA 479  
QY 302 ATATTTGGCTAATGAGGAAGCAAT 326

```

Db 480 AGATTGGTTGATGAGAACACAT 504
|||||
RESULT 3
AB067779
LOCUS
DEFINITION Tamias sibiricus gene for HP-20, complete cds.
ACCESSION AB067779
VERSION AB067779.1 GI:15706341
KEYWORDS
SOURCE
ORGANISM
TAMIAS sibiricus DNA, clone:lambdaCM20G1.
TAMIAS sibiricus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Scurinae;
TAMIAS.
REFERENCE
1 Ono,M., Hosoe,Y., Azuma,S., Shoji,M., Nara,K., Kondo,N., Shiba,T.
and Takamatsu,N.
HNF-1 regulates the liver-specific transcription of the chipmunk
HP-20 gene
Gene 277 (1-2), 121-127 (2001)
21488336
2 (bases 1 to 5121)
Onc.M.
Direct Submission
Submitted (04-AUG-2001) Motoharu Ono, Kitasato University,
Department of Biosciences, School of Science; 1-15-1 Kitasato,
Sagamihara, Kanagawa 228-8555, Japan
(E-mail:ms9805mestu.sci.kitasato-u.ac.jp, Tel:81427789408,
Fax:81427789408)
Location/Qualifiers
1. .5121
/organism="Tamias sibiricus"
/db_xref="taxon:64680"
/clone="lambdaCM20G1"
/notes="synonym:Tamias asiaticus"
1518. .1684
join(1678..1684,2573..2710,3968..4413)
/gene="HP-20"
CDS
join(1678..1684,2573..2710,3968..4413)
/gene="HP-20"
/codon_start=1
/product="HP-20"
/db_xref="GI:15706342"
/translation="MTDAWRLAIFVLMVNLNDQVSCSGPPGVPYGVPGVPGPRCP
PQOPGAAGRPDGPGRKPSVKPCRSFTVKFSGRLPPPEVPVFTVLVNTQKDL
KASTGVNCEVPEGNHFSFDVELYHCKVKIGLMKNHIOVMEXHQLSKNEYENASGAMI
MPLROGDKVWLEADVETERPDAQVYVFSGLISS"
exon
3968..5000
/gene="HP-20"
exon
1571 a 1018 c 1090 g 1442 t
BASE COUNT
ORIGIN
Query Match 35.2%; Score 139.2; DB 10; Length 5121;
Best Local Similarity 71.6%; Pred. No. 9.9e-31;
Matches 199; Conservative 0; Mismatches 73; Indels 6; Gaps 1;
QY 49 GGTGCACAGGTTTACCACAATATACAGGAGAAATAGTGAATGACAAATGCCCTGT 108
|||||
Db 3967 GCGCTCGGGTAGGCAGGAGACCCAGACCAAAAGGGCCCGATGTAATGCCCTG- 4025
|||||
QY 109 CCTGATATAGAAAGTCAGCCTTTACTGTGAAGCTCAGTGGAAACTTCCTCTTCCTTC 168
|||||
Db 4026 ----CAGAGAGAGGTGACGCTTTCACGGTGAAGTTTACGCGGAAGGCTCCCTCCACCTTCG 4080
|||||
QY 169 AAGCCCATCATCTTACAGGGGTCCTGTACAATGCCAGAGGGATTTAAAGAGAGGCCAT 228
|||||
Db 4081 GAGCCTGTGCTCTACAGAGGTCCTGTACATACCAGAGGAGACTTGAAGGAGGACACA 4140
|||||
QY 229 GGAGTCTTTGCTGCGAGGGTGCCTGGGAATTACTACTCCAGCTTTGATGTTGAGCTGCAT 288
|||||

Db 4141 GGAGTCTTTAACTGCGTGAGCGCTGGAAATACCATTTACGTTGATGTTGAGCTTTAC 4200
|||||
QY 289 CATTGCAAGGTGAATATTTGCTAATAGCAAGCAAT 326
|||||
Db 4201 CACTGCAAGGTGAAGATTGGTTGATGAAGAACACAT 4238
|||||

RESULT 4
TMSHP25
LOCUS
DEFINITION Tamias asiaticus mRNA for HP-25, complete cds.
ACCESSION D12975
VERSION D12975.1 GI:287469
KEYWORDS
HP-25; collagen-like domain; hibernation-related protein; plasma
protein.
SOURCE
ORGANISM
TAMIAS asiaticus liver cDNA to mRNA, clone:pcM25-3.
TAMIAS sibiricus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Scurinae;
TAMIAS.
REFERENCE
1 (bases 1 to 2005)
Takamatsu,N., Ohba,K., Kondo,J., Kondo,N. and Shiba,T.
Hibernation-associated gene regulation of plasma proteins with a
collagen-like domain in mammalian hibernators
Mol. Cell. Biol. 13 (3), 1516-1521 (1993)
93180798
2 (bases 1 to 2005)
Takamatsu,N.
Direct Submission
Submitted (21-AUG-1992) Nobuhiko Takamatsu, School of Hygienic
Sciences, Kitasato University, Laboratory of Molecular Biology;
1-15-1 Kitasato, Sagamihara, Kanagawa 228, Japan (tel:0427-78-9408,
Fax:0427-78-9403)
Location/Qualifiers
1. .2005
/organism="Tamias sibiricus"
/db_xref="taxon:64680"
/clone="pcM25-3"
/tissue_type="liver"
114..761
/notes="collagen-like domain at nt 231-356"
/codon_start=1
/product="HP-25"
/db_xref="GI:287470"
/translation="NPAQGGALSMGAAGFWILVLSITLADSNNGNSEPCGPPGP
PQPPGIPGPGAPGALGPPGPGVPGIPGPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP
PFQPIVFKKALYNQEGHFNATGEFSCVLPVNFQVDFDLRLFOSSVKIRMRDGIQVR
EKEAQAQNDYSYKHAMGVSIMALGKDKVWLESKLKGTSESEKGIHIVFFGYLLYK"
sig_peptide
114..197
mat_peptide
198..758
/polyA_signal
1983..1988
/polyA_site
2005
BASE COUNT 569 a 472 c 494 g 470 t
ORIGIN
Query Match 21.0%; Score 83; DB 10; Length 2005;
Best Local Similarity 54.9%; Pred. No. 7.6e-14;
Matches 186; Conservative 0; Mismatches 150; Indels 3; Gaps 1;
QY 10 CCAGCACACCCAGGCCCCAGAGAAGTGGGCCTCTGTCGCCACCAAGTTTACCACAA 69
|||||
Db 264 CCAGCTTCCAGCGCTCTCTGGAGCACTTGTCCACAGAGACCCCGAGGTGCCAGGA 323
|||||
QY 70 TATACAGGAGAAATAGTGAATGACAAA---ATGCCCTCTGCTGATATAGAAAGTCA 126
|||||
Db 324 ATACCAGGACCAACAAAGTCTCTCTGGAGAGCTTGAGAAGTGCTCATCGCGACCAATCT 383
|||||
QY 127 GCCTTTACTGTGAAGCTCAGTGGNAACACTTCCTCTCTCTTTCAGAGCCCATCTTCACA 186
|||||
Db 384 GCCTTTGCTGAGCTGAGTGAACGGCCCCCAGAGAGCCCTTCCAGGCCCATGCTTCAAG 443
|||||
```

```
QY 187 GGGTCTCTGTACAAATGCCAGAGGATTTAAAGAGGCCATGGGAGTCTTTGCTTGCAGG 246
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 444 GAGGCTCTGTACAATCAGGAGGCCACTTCAACATGGCCACGGAGAGTTCAGCTGTGTC 503
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 247 GTGCTGCGAATTTACTACTCAGCTTTTCATGTTGAGCTGCATCATTCGAAGGTGAATTT 306
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 504 CTCCTCTGTGTGACAAATTTGGCTTTTGACATTCGACATTCAGATTCCTGTGAAGATC 563
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 307 TGGCTAATGAGGAACAAATTTTGGCTAATAAGGAAGAA 345
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 564 AGGCTCATGAGGATGGTATCCAGGTCAGAGAGAAGAA 602
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
LOCUS AB067813s3 559 bp DNA linear ROD 18-JAN-2002
DEFINITION Callosciurus caniceps HP-25 pseudogene, exon 3, complete cds.
ACCESSION AB069679
VERSION AB069679.1 GI:18149913
SEGMENT 3 of 3
KEYWORDS Callosciurus caniceps DNA, clone:lambda TS25G1.
SOURCE Callosciurus caniceps
ORGANISM Callosciurus caniceps
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
Callosciurus.
REFERENCE 1
AUTHORS Kojima,M., Shiba,T., Kondo,N. and Takamatsu,N.
TITLE The tree squirrel HP-25 gene is a pseudogene
JOURNAL Eur. J. Biochem. 268 (22), 5997-6002 (2001)
MEDLINE 21579740
REFERENCE 2 (bases 1 to 559)
AUTHORS Takamatsu,N.
TITLE Direct Submission
JOURNAL Submitted (06-AUG-2001) Nobuhiko Takamatsu, Kitasato University,
Department of Biosciences, School of Science; 1-15-1 Kitasato,
Sagamihara, Kanagawa 228-8555, Japan
(E-mail:takamatu@jet.sci.kitasato-u.ac.jp, Tel:81427789408,
Fax:81427789408)

FEATURES
    source
        1..559
            /organism="Callosciurus caniceps"
            /db_xref="taxon:64664"
            /clone="lambda TS25G1"
            /join(AB067813.1:1503..1658,AB069678.1:26..185,39..559)
            /gene="HP-25"
            /gene="HP-25"
            /join(AB067813.1:1503..1658,AB069678.1:26..185,39..496)
            /pseudogene="HP-25"
            /codon_start=1
            /product="HP-25"
            39..>559
            /gene="HP-25"
            /number=3
            /pseudo

BASE COUNT 150 a 128 c 141 g 140 t
ORIGIN
Query Match 17.3%; Score 68.6; DB 10; Length 559;
Best Local Similarity 55.5%; Pred. No. 1.5e-09;
Matches 156; Conservative 0; Mismatches 119; Indels 6; Gaps 1;

QY 40 GGGCTCTCTGTGTCACAGGTTTACCACAATATACAGAGAAATAGTGAATGACAAAA 99
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38 GGACATCCAGGTATACAGGAGGCCAGGATGACAGGCCCTCTGGAGACATGAGACG 97
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 100 TGCCCTCTCTGTATAGAAAGGTCAGCCTTTACTGTGAAGCTCAGTGGAAAACTCTCT 159
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 98 TGCCCATCACC-----ACCAAAATCTGCCCTTGGCGTGAAGCTGAGCGTCCCCCA 151
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 160 CTTCTTTCAAGCCCATCTTTCACAGGGGTCTCTGTACAATGCCAGGAGGATTTAAG 219
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 152 GGGTCTCTCCAGCCCATCATATTTCAAGGAGTCTCTGTACAAACGGGAGACCACATTAAC 211
QY 220 GAGGCGATGGGAGTCTTTGCTTGCAGGGTGCCTCGGGAATTTACTACTCCAGCTTTGATGTT 279
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 212 GTGACCACCTGGAGAGTTTCAGCTATACCAACCCCGGTGTGTACAAATTTTGGCTTTGACATT 271
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 280 GAGCTGCATCATTCGCAAGGTGAATATTTGGCTTAATGAGGAA 320
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 272 GGACTGTTTCAGAGTCTCTGTAAGATAAGTCTCATGAGGAA 312
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
HS302D9/c 145880 bp DNA linear PRI 12-DEC-1999
LOCUS Human DNA sequence from clone RPI-302D9 on chromosome 22 Contains
DEFINITION GSSs, complete sequence.
ACCESSION 282198
VERSION 282198.2 GI:6572207
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 145880)
AUTHORS Bridgeman,A.
TITLE Direct Submission
JOURNAL Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Dec 13, 1999 this sequence version replaced gi:3164067.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C/elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
RPI-302D9 is from the library RPCI-1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
This sequence is the entire insert of clone RPI-302D9 The true left
end of clone CTA-282F2 is at 69682 in this sequence. The true right
end of clone CTA-41562 is at 55167 in this sequence.

FEATURES
    source
        1..145880
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="22"
            /clone="RPI-302D9"
            /clone_lib="RPCI-1"
            188..245
            repeat_region
                /note="MER3 repeat: matches 144..209 of consensus"
            repeat_region
                /note="AluX repeat: matches 1..312 of consensus"
            repeat_region
                /note="MER3 repeat: matches 1..144 of consensus"
            repeat_region
                /note="MER5A repeat: matches 26..187 of consensus"
```

```
repeat_region 1033. .1336
/note="AluSp repeat: matches 1. .299 of consensus"
repeat_region 1450. .1583
/note="MIR repeat: matches 24. .160 of consensus"
repeat_region 1687. .1752
/note="L2 repeat: matches 2593. .2661 of consensus"
repeat_region 2350. .2660
/note="AluSc repeat: matches 3. .309 of consensus"
repeat_region 2684. .2981
/note="AluSg repeat: matches 2. .300 of consensus"
repeat_region 3323. .3343
/note="MLTIE repeat: matches 116. .136 of consensus"
repeat_region 3344. .3652
/note="AluY repeat: matches 1. .309 of consensus"
repeat_region 3653. .3928
/note="MLTIE repeat: matches 136. .359 of consensus"
repeat_region 3929. .4278
/note="THEIB repeat: matches 3. .364 of consensus"
repeat_region 4279. .4485
/note="MLTIE repeat: matches 359. .568 of consensus"
repeat_region 5073. .5176
/note="52 copies 2 mer ct 78 conserved"
repeat_region 5181. .5491
/note="AluDb repeat: matches 1. .311 of consensus"
repeat_region 6369. .6485
/note="L2 repeat: matches 2579. .2705 of consensus"
repeat_region 6647. .6685
/note="WADEL repeat: matches 1. .23 of consensus"
repeat_region 6686. .6987
/note="AluSx repeat: matches 1. .302 of consensus"
repeat_region 6988. .7036
/note="WADEL repeat: matches 23. .77 of consensus"
repeat_region 7482. .7754
/note="AluJb repeat: matches 9. .290 of consensus"
repeat_region 7775. .8060
/note="AluDu repeat: matches 1. .295 of consensus"
repeat_region 8414. .8551
/note="L2 repeat: matches 2553. .2706 of consensus"
repeat_region 8914. .9030
/note="MIR repeat: matches 147. .262 of consensus"
repeat_region 9110. .9280
/note="MIR repeat: matches 91. .262 of consensus"
repeat_region 9283. .9412
/note="MIR repeat: matches 15. .144 of consensus"
repeat_region 9521. .9679
/note="FAM repeat: matches 3. .161 of consensus"
repeat_region 9820. .10225
/note="WSTB repeat: matches 2. .425 of consensus"
misc_feature complement(10179. .10678)
/note="match: GSS: Em:B56592"
misc_feature complement(10204. .10728)
/note="match: GSS: Em:AQ701486"
misc_feature complement(10249. .10706)
/note="match: GSS: Em:AQ225495"
repeat_region 10312. .10383
/note="MIR repeat: matches 79. .150 of consensus"
misc_feature 10718. .11310
/note="match: GSS: Em:B14024"
misc_feature 10784. .11201
/note="match: GSS: Em:B43656"
repeat_region 11838. .11946
/note="MIR repeat: matches 20. .137 of consensus"
repeat_region 12174. .12445
/note="L2 repeat: matches 1988. .2275 of consensus"
repeat_region 12444. .12642
/note="MIR repeat: matches 63. .241 of consensus"
misc_feature 13017. .13369
/note="match: STS: Em:G49301"
repeat_region 13331. .13397
/note="MIR repeat: matches 174. .244 of consensus"
repeat_region 13398. .13698
/note="AluSp repeat: matches 1. .302 of consensus"
repeat_region 13699. .13810
/note="MIR repeat: matches 76. .174 of consensus"
13806. .13919
/note="MIR repeat: matches 77. .189 of consensus"
13945. .14060
/note="MIR repeat: matches 24. .142 of consensus"
14061. .14367
/note="AluY repeat: matches 1. .301 of consensus"
14368. .14452
/note="MIR repeat: matches 141. .225 of consensus"
14589. .14679
/note="MIR repeat: matches 173. .262 of consensus"
14597. .15201
/note="match: GSS: Em:AQ553482"
14616. .15060
/note="match: GSS: Em:AQ370601"
14868. .15040
/note="MIR repeat: matches 49. .233 of consensus"
15071. .15188
/note="L2 repeat: matches 2112. .2239 of consensus"
15304. .15399
/note="MLTIB repeat: matches 1. .99 of consensus"
15490. .15662
/note="AluSgl repeat: matches 2. .114 of consensus"
15669. .15727
/note="MLTIB repeat: matches 119. .178 of consensus"
15728. .16027
/note="AluSc repeat: matches 1. .299 of consensus"
16028. .16245
/note="MLTIB repeat: matches 178. .390 of consensus"
16546. .16854
/note="AluY repeat: matches 1. .300 of consensus"
18296. .18323
/note="WSTA repeat: matches 2. .29 of consensus"
18324. .18392
/note="MER66-internal repeat: matches 4919. .4993 of consensus"
18393. .18712
/note="AluJb repeat: matches 1. .311 of consensus"
18713. .19133
/note="MER66-internal repeat: matches 4548. .4919 of consensus"
misc_feature complement(18872. .19230)
/note="match: GSS: Em:AQ005063"
misc_feature 19251. .19719
/note="match: GSS: Em:B14179"
repeat_region 19537. .20290
/note="HERVPH21 repeat: matches 4657. .5784 of consensus"
repeat_region 20317. .20382
/note="33 copies 2 mer ta 68 conserved"
repeat_region 20513. .20666
/note="77 copies 2 mer tt 70 conserved"
repeat_region 20682. .21008
/note="AluSgl repeat: matches 1. .306 of consensus"
repeat_region 21239. .21553
/note="HUERS-P3 repeat: matches 4410. .4713 of consensus"
repeat_region 21882. .22254
/note="THEIB repeat: matches 1. .364 of consensus"
repeat_region 22302. .22537
/note="MER66-internal repeat: matches 2186. .2417 of consensus"
repeat_region 22538. .22850
/note="AluSp repeat: matches 1. .313 of consensus"
repeat_region 22851. .23801
/note="MER66-internal repeat: matches 1210. .2186 of consensus"
repeat_region 23905. .23989
/note="MER66-internal repeat: matches 3017. .3102 of consensus"
Query Match 17.2%; Score 68.2; DB 9; Length 145880;
Best Local Similarity 58.7%; Pred. No. 3e-09;
Matches 118; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
```

```
Db 90879 AGATCTGATTTGGTGAAGTACTGATCGAGCCCTCCAGTGCCCTTAGCCCATTTGT 90820
Qy 180 CTTACAGGGTCTGTACAACTGCCAGAGGATTTAAAGGAGGCCATGGGAGTCTTTTCG 239
Db 90819 CTTCAAGGAAGCCCTTTATAAATACACAGTTCATTCATTTCTCTGGTGAAGGATTAC 90760
Qy 240 TTGCAGGGTCCCTGGGAATTAATACTACTCCAGCTTTGATGTTGAGTCGATCATTTGCAAGT 299
Db 90759 CTGTACCACTCCCTAGTGTGACACTCTGTTTGAATTTGAGCTGTTTCAGAGCTGTGC 90700
Qy 300 GAATATTTGGCTAATCAGGAA 320
Db 90699 AATGTGGGTATTATGAGGAA 90679

RESULT 7
LOCUS H5732E4
DEFINITION Human DNA sequence from clone CTA-732E4 on chromosome 22q12.1
ACCESSION AL008722
VERSION AL008722.16 GI:5419644
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 90497)
Lloyd.D.
Direct Submission
Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 8, 1999 this sequence version replaced gi:5304865.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sv., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C-elegans/wormpep
from the human BAC library described in U-J. Kim et al. (1996)
Genomics 34, 213-218.
VECTOR: pBelobAC11
This sequence is the entire insert of clone CTA-732E4 The true left
end of clone RPL1-541J16 is at 5510 in this sequence. The true
right end of clone CTA-544A11 is at 41939 in this sequence.
FEATURES
source
1. 90497
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/map="q12.1"
/clone="CTA-732E4"
/clone.lib="CIT978SK-A2"
repeat_region 2..631
/note="L1M1 repeat: matches 994..1640 of consensus"
repeat_region 632..757
/note="AluYb8 repeat: matches 1..306 of consensus"
```

```
/note="AluJb repeat: matches 1..126 of consensus"
758..1385
/note="L1P repeat: matches 3548..4173 of consensus"
1374..1869
/note="L1P repeat: matches 3221..3716 of consensus"
1952..2076
/note="L1MD2 repeat: matches 5142..5259 of consensus"
2760..3005
/note="L1MD2 repeat: matches 5259..6173 of consensus"
3418..3477
/note="30 copies 2 mer ac 90 conserved"
3423..3478
/note="14 copies 4 mer caca 94 conserved"
3670..3742
/note="MER82 repeat: matches 580..653 of consensus"
3763..4128
/note="MER82 repeat: matches 1..386 of consensus"
3879..4353
/note="match: GSS: Em:AQ559531"
4364..4569
/note="L12 repeat: matches 2535..2750 of consensus"
4612..4930
/note="AluJo repeat: matches 1..307 of consensus"
4934..5220
/note="AluX repeat: matches 1..287 of consensus"
5221..5268
/note="16 copies 3 mer taa 97 conserved"
complement(5266..5277)
/note="match: GSS: Em:AQ028870"
complement(5282..5740)
/note="match: GSS: Em:AQ557814"
5417..5569
/note="MIR repeat: matches 2..167 of consensus"
5645..5708
/note="32 copies 2 mer at 71 conserved"
6469..6726
/note="match: STS: Em:G03694; match: STS: Em:G03694"
7573..7749
/note="MIR repeat: matches 22..207 of consensus"
10322..16495
/note="L1PA2 repeat: matches 7..6146 of consensus"
16265..16787
/note="match: GSS: Em:AQ186300"
16926..16967
/note="21 copies 2 mer aa 76 conserved"
17850..17971
/note="AluJo repeat: matches 19..147 of consensus"
17972..18285
/note="AluSg1 repeat: matches 1..309 of consensus"
18286..18472
/note="AluJo repeat: matches 147..309 of consensus"
19710..20005
/note="AluX repeat: matches 1..296 of consensus"
20789..21262
/note="match: GSS: Em:BL4443"
20811..21058
/note="match: STS: Em:L04555"
21369..21551
/note="MIR repeat: matches 2..187 of consensus"
22494..22697
/note="MIR repeat: matches 35..236 of consensus"
23443..23755
/note="AluX repeat: matches 1..312 of consensus"
24052..24348
/note="AluSg repeat: matches 1..295 of consensus"
25623..25670
/note="24 copies 2 mer ac 77 conserved"
26003..26192
/note="MIR repeat: matches 20..216 of consensus"
27067..27138
/note="Charlie4 repeat: matches 1902..1961 of consensus"
27363..27661
/note="AluYb8 repeat: matches 1..306 of consensus"
```



[illegible]



```

/note="assembly_fragment"
48699. .52567
/note="assembly_fragment"
52668. .55272
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
55373. .59485
/note="assembly_fragment"
59586. .63994
/note="assembly_fragment"

```

```

Query Match      16.0%; Score 63.2; DB 2; Length 144514;
Best Local Similarity 74.1%; Pred. No. 9.7e-08;
Matches 80; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

```

```

QY 289 CATTCAAGGTGANTATTGGCTAATGAGGAGCAAAATTTTGGCTAATAAGGAAGAAATTT 348
|||||
Db 133169 CCTAGAGATTGTGAACTTTGAACCTGGAAGAGATGATTTAGAGTATCAGGGAAGAAATTT 133228
|||||
QY 349 TCTAAGCAGCAAGCATTTCAAGAGGTGACTTGGGTGCTCTTAAAGGCA 396
|||||
Db 133229 TCTAAGCAGCAAGCATTTCAAGAGGTGACTTGGGTGCTCTTAAAGGCA 133276
|||||

```

```

RESULT 11
AC025112      149038 bp      DNA      linear      HTG 08-APR-2000
LOCUS
DEFINITION
Homo sapiens chromosome 6 clone RP11-608N7 map 6, WORKING DRAFT
SEQUENCE, 27 unordered pieces.
AC025112
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 149038)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 6, clone RP11-608N7
Unpublished
2 (bases 1 to 149038)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bida,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collumore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferrelle,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tefaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 8, 2000 this sequence version replaced gi:7158923.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

```

```

TITLE
JOURNAL
COMMENT

```

```

Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6882
Center clone name: 608_N7
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 132985 bases at least Q40
Consensus quality: 140748 bases at least Q30
Consensus quality: 143894 bases at least Q20
Insert size: 138000; agarose-fp
Insert size: 146438; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 3.4 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 165: contig of 165 bp in length
166 265: gap of 100 bp
266 1355: contig of 1050 bp in length
1356 1455: gap of 100 bp
1456 1752: contig of 297 bp in length
1753 1852: gap of 100 bp
1853 3336: contig of 1484 bp in length
3337 3436: gap of 100 bp
3437 4814: contig of 1378 bp in length
4815 4914: gap of 100 bp
4915 6853: contig of 1939 bp in length
6854 6953: gap of 100 bp
6954 9114: contig of 2161 bp in length
9115 9214: gap of 100 bp
9215 12071: contig of 2857 bp in length
12072 12171: gap of 100 bp
12172 14633: contig of 2462 bp in length
14634 14733: gap of 100 bp
14734 17354: contig of 2621 bp in length
17355 17454: gap of 100 bp
17455 21910: contig of 4456 bp in length
21911 22010: gap of 100 bp
22011 26287: contig of 4277 bp in length
26288 26387: gap of 100 bp
26388 31094: contig of 4707 bp in length
31095 31194: gap of 100 bp
31195 35408: contig of 4214 bp in length
35409 35508: gap of 100 bp
35509 40764: contig of 5256 bp in length
40765 40864: gap of 100 bp
40865 45639: contig of 4775 bp in length
45640 45739: gap of 100 bp
45740 50826: contig of 5087 bp in length
50827 50926: gap of 100 bp
50927 56617: contig of 5691 bp in length
56618 56717: gap of 100 bp
56718 63406: contig of 6689 bp in length
63407 63506: gap of 100 bp
63507 71321: contig of 7815 bp in length
71322 71421: gap of 100 bp
71422 81641: contig of 10220 bp in length
81642 81741: gap of 100 bp
81742 91982: contig of 10241 bp in length
91983 92082: gap of 100 bp
92083 102357: contig of 10275 bp in length
102358 102457: gap of 100 bp
102458 112952: contig of 10495 bp in length
112953 113052: gap of 100 bp

```

\* 113053 124237: contig of 11185 bp in length  
\* 124238 124337: gap of 100 bp  
\* 124338 135694: contig of 11357 bp in length  
\* 135695 135794: gap of 100 bp  
\* 135795 149038: contig of 13244 bp in length.

## FEATURES

```
source
1. .149038
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6"
/clone="RP11-608N7"
/clone_lib="RPC1-11 Human Male BAC"
1. .165
/note="assembly_fragment
clone_end:T7
vector_side:right"
misc_feature
266. .1355
/note="assembly_fragment"
misc_feature
1456. .1752
/note="assembly_fragment
clone_end:SP6
vector_side:left"
1853. .3336
/note="assembly_fragment"
misc_feature
3437. .4814
/note="assembly_fragment"
misc_feature
4915. .6853
/note="assembly_fragment"
6354. .9114
/note="assembly_fragment"
9215. .12071
/note="assembly_fragment"
12172. .14633
/note="assembly_fragment"
14734. .17354
/note="assembly_fragment"
17455. .21910
/note="assembly_fragment"
22011. .26287
/note="assembly_fragment"
26388. .31094
/note="assembly_fragment"
31195. .35408
/note="assembly_fragment"
35509. .40764
/note="assembly_fragment"
40865. .45639
/note="assembly_fragment"
45740. .50826
/note="assembly_fragment"
50927. .56617
/note="assembly_fragment"
56718. .63406
/note="assembly_fragment"
63507. .71321
/note="assembly_fragment"
71422. .81641
/note="assembly_fragment"
81742. .91982
/note="assembly_fragment"
92083. .102357
/note="assembly_fragment"
102458. .112952
/note="assembly_fragment"
113053. .124237
/note="assembly_fragment"
124338. .135694
/note="assembly_fragment"
135795. .149038
/note="assembly_fragment"
44765 a 26643 c 27220 g 47805 t 2605 others
BASE COUNT
ORIGIN
```

```
Query Match          16.0%; Score 63.2; DB 2; Length 149038;
Best Local Similarity 74.1%; Pred. No. 9.7e-08;
Matches 80; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 289 CATTGCAAGGTGAATATTGGCTAATGAGGAAGCAAAATTTGGCTAATAGGAAGAATT 348
Db 90741 CCTAGAGATTGTGAACCTTGAACCTGAAAGAGATGATGATAGATGACGAGGAAGAATT 90800

QY 349 TCTAAGCAGCAAAAGCATTCAAGAGGTGACATTGGTGGCTGTTAAAGGCA 396
Db 90801 TCTAAGCAGCAAAAGCATTCAAGAGGTGACATTGGTGGCTGTTAAAGGCA 90848
```

RESULT 12  
AC015553/CLOCUS  
AC015553/1DEFINITION  
Homo sapiens 9 BAC RP11-100N10 (Roswell Park Cancer Institute Human BAC Library) complete sequence.ACCESSION  
AC015553VERSION  
AC015553.21 GI:13489133KEYWORDS  
HTG.SOURCE  
Homo sapiens.ORGANISM  
Homo sapiensREFERENCE  
1 (bases 1 to 162063)AUTHORS  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Lucier,A., Lucier,R., Luna,R., Louisaged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., and Gibbs,R.TITLE  
Direct SubmissionJOURNAL  
UnpublishedREFERENCE  
2 (bases 1 to 162063)AUTHORS  
Worley,K.C.TITLE  
Direct SubmissionJOURNAL  
Submitted (17-NOV-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One



```
Matches 74; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 304 ATTGGCTAATGAGGAAGCAAAATTTGGCTAATAAGGAAGCAAAATTTCTAAGCAGCAAGC 363
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29373 ACTTTGAACCTTGAGAGATGATTTAGGCTATGCGAAGAAATTTCTAAGCAGCAAGC 29314
QY 364 ATTCAGAGGTGACTGGTGGTCTGCTTTAAAGGCA 396
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29313 ATTCAGAGGTGACTGGTGGTCTGCTTTAAAGACA 29281

RESULT 13
AL161623/c
LOCUS
DEFINITION
  Homo sapiens chromosome 9 clone RP11-IK21, linear HTG 10-JUL-2001
  PROGRESS ***, 20 unordered pieces.
ACCESSION
  AL161623
VERSION
  12 GI:9930848
KEYWORDS
  HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  McElay, K.
  Direct Submission
  Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
  CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
  Requests: clonerequest@sanger.ac.uk
  On Aug 27, 2000 this sequence version replaced gi:9863606.
  ----- Genome Center
  Center: Sanger Centre
  Center code: SC
  Web site: http://www.sanger.ac.uk
  Contact: humquery@sanger.ac.uk
  ----- Project Information
  Center project name: bALK21
  ----- Summary Statistics
  Assembly program: XGAP4; version 4.5
  Sequencing vector: plasmid; L08752; 100% of reads
  Chemistry: Dye-terminator ET-amersham; 24% of reads
  Dye-terminator Big Dye; 75% of reads
  Consensus quality: 164370 bases at least Q40
  Consensus quality: 168213 bases at least Q30
  Consensus quality: 170527 bases at least Q20
  Insert size: 173234; sum-of-contigs
  Insert size: 168172; 4.4% error; agarose-fp
  Quality coverage: 4.13x in Q20 bases; sum-of-contigs Quality
  coverage: 4.57x in Q20 bases; agarose-fp
  -----
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 20 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
  *
  * 1 3511: contig of 3511 bp in length
  * 3512 3611: gap of 100 bp
  * 3612 25895: contig of 22284 bp in length
  * 25896 25995: gap of 100 bp
  * 25996 31543: contig of 5548 bp in length
  * 31544 31643: gap of 100 bp
  * 31644 39248: contig of 7605 bp in length
  * 39249 39348: gap of 100 bp
  * 39349 49948: contig of 10600 bp in length
  * 49949 50048: gap of 100 bp
  * 50049 66131: contig of 16083 bp in length
  * 66132 66231: gap of 100 bp
  * 66232 80131: contig of 13900 bp in length
  * 80132 80231: gap of 100 bp
  * 80232 89785: contig of 9554 bp in length

89786 89885: gap of 100 bp
89886 92978: contig of 3093 bp in length
92979 93078: gap of 100 bp
93079 97528: contig of 4450 bp in length
97529 97628: gap of 100 bp
97629 101520: contig of 3892 bp in length
101521 101620: gap of 100 bp
101621 112966: contig of 11346 bp in length
112967 113066: gap of 100 bp
113067 119217: contig of 6151 bp in length
119218 119317: gap of 100 bp
119318 121556: contig of 2239 bp in length
121557 121656: gap of 100 bp
121657 134400: contig of 12744 bp in length
134401 134500: gap of 100 bp
134501 144934: contig of 10434 bp in length
144935 145034: gap of 100 bp
145035 148818: contig of 3784 bp in length
148819 148918: gap of 100 bp
148919 152178: contig of 3260 bp in length
152179 152278: gap of 100 bp
152279 165020: contig of 12742 bp in length
165021 165120: gap of 100 bp
165121 175134: contig of 10014 bp in length.

FEATURES
      source
      1..175134
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="9"
      /clone="RP11-IK21"
      /clone_lib="RPC1-11.1"
      1..3511
      /note="assembly_fragment:00168"
      fragment_chain:1
      3612..25895
      /note="assembly_fragment:00194"
      fragment_chain:1
      25996..31543
      /note="assembly_fragment:00852"
      fragment_chain:2
      31644..39248
      /note="assembly_fragment:00602"
      fragment_chain:2
      39349..49948
      /note="assembly_fragment:01593"
      fragment_chain:3
      50049..66131
      /note="assembly_fragment:01966"
      fragment_chain:3
      66232..80131
      /note="assembly_fragment:02033"
      fragment_chain:4
      80232..89785
      /note="assembly_fragment:00372"
      fragment_chain:4
      89886..92978
      /note="assembly_fragment:02305"
      fragment_chain:5
      93079..97528
      /note="assembly_fragment:00011"
      fragment_chain:5
      97629..101520
      /note="assembly_fragment:00955"
      101621..112966
      /note="assembly_fragment:01248"
      113067..119217
      /note="assembly_fragment:01397"
      119318..121556
      /note="assembly_fragment:01523"
      121657..134400
      /note="assembly_fragment:01568"
      134501..144934
      /note="assembly_fragment:01617"
```



```
mat_peptide      155..709
                  /product="HP-27"
polyA_signal     1369..1374
polyA_site       1385
BASE COUNT      439 a      285 c      311 g      350 t
ORIGIN
Query Match      15.8%; Score 62.4; DB 10; Length 1385;
Best Local Similarity 55.6%; Pred. No. 1.2e-07;
Matches 120; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
QY 105 CTGCTCGATATAGAAAGGTCAGCCCTTACTGTGAAGCTCAGTGGAAAACCTTCCTCTTCC 164
Db      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
313 CTGCCACAGCAAAAGGAACATCGGCCCTTTCAGTGAAGGCAATGAGTGGCCCCCAGCTCC 372
QY      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
165 TTTCAGGCCCATCATCTTCACAGGGTCTGTACAAATGCCCCAGAGGGATTTAAAGGAGGC 224
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
373 CTCCCAGCCCGTATCTTCAAGGAAGCCCTGCATGACGCTCAGGGACACTTTGATCTGGC 432
QY      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
225 CATGGGAGTCTTTGCTTGCAAGGTGCCTGGGAATTACTACCAAGCTTTTGATGTTGAGCT 284
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
433 CACTGGTGTGTTTCACCTGCCCCAGTCCCGAGACTCTACCAGTTTGGATTTCACATTGAAGC 492
QY      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
285 GCATCATTGCAAGGTGAATATTTGGCTAATGAGGAA 320
Db      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
493 TGTCAGAGGGCTGTGAAGGTGAGCCTCATGAGAAA 528
```

Search completed: February 20, 2003, 04:48:41  
Job time : 2075.56 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:48:42 ; Search time 262.925 Seconds  
(without alignments)  
6406.126 Million cell updates/sec

Title: US-09-997-610-1\_COPY\_53\_156

Perfect score: 104

Sequence: 1 gctggacctccagcacacc.....agtgaaatgacaaatgcc 104

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: gb\_gss:\*
- 18: em\_gss\_hum:\*
- 19: em\_gss\_inv:\*
- 20: em\_gss\_pln:\*
- 21: em\_gss\_vrt:\*
- 22: em\_gss\_fun:\*
- 23: em\_gss\_mam:\*
- 24: em\_gss\_mus:\*
- 25: em\_gss\_other:\*
- 26: em\_gss\_pro:\*
- 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
c					
1	38.4	36.9	531	10	BE121287
2	38	36.5	483	13	BI322379
3	38	36.5	552	13	BI863711
4	37.4	36.0	889	14	BQ939199
5	36.8	35.4	695	10	BE285930
6	36.8	35.4	776	13	BI554667

7	36.2	34.8	360	14	D69730
c					
8	36.2	34.8	632	13	BJ136441
9	36.2	34.8	726	13	BJ108893
10	36	34.6	461	13	BI323061
11	35.8	34.4	675	12	BG310518
12	35.4	34.0	287	10	BE029138
13	35.2	33.8	468	10	BE580165
14	35.2	33.8	483	10	BE579244
15	35.2	33.8	814	10	BE286914
16	34.8	33.5	613	10	AW179928
17	34.6	33.3	300	14	D70222
18	34.6	33.3	360	14	D69877
19	34.6	33.3	407	10	BE581103
20	34.6	33.3	420	10	BE579707
21	34.6	33.3	441	10	BE581944
22	34.6	33.3	478	10	BE579460
23	34.6	33.3	491	10	BE579541
24	34.6	33.3	492	10	BE579271
25	34.6	33.3	496	10	BE579572
26	34.6	33.3	525	10	AW675997
27	34.6	33.3	575	10	AW179710
28	34.6	33.3	585	10	AW179840
c					
29	34.6	33.3	601	9	AU217866
30	34.6	33.3	605	10	AW347957
31	34.6	33.3	624	12	BE758489
32	34.6	33.3	640	10	AW179745
c					
33	34.6	33.3	645	13	BJ139557
c					
34	34.6	33.3	647	9	AU220646
35	34.6	33.3	647	10	AW409404
36	34.6	33.3	647	10	AW409406
37	34.6	33.3	648	10	AW874708
38	34.6	33.3	654	10	AW348036
c					
39	34.6	33.3	666	13	BJ144812
c					
40	34.6	33.3	672	13	BJ138817
c					
41	34.6	33.3	672	13	BJ140783
c					
42	34.6	33.3	673	13	BJ144467
c					
43	34.6	33.3	674	13	BJ146794
c					
44	34.6	33.3	693	9	AU217833
45	34.6	33.3	701	10	AW179849

ALIGNMENTS

RESULT 1	BE121287/c	BE121287	531 bp	mRNA	linear	EST 13-JUN-2000
LOCUS	UI-R-CA0-baw-b-08-0-UI.s1	UI-R-CA0	Rattus norvegicus	CDNA	clone	
DEFINITION	UI-R-CA0-baw-b-08-0-UI 3', mRNA sequence.					
ACCESSION	BE121287					
VERSION	BE121287.1	GI:8513392				
KEYWORDS	EST.					
SOURCE	Norway rat.					
ORGANISM	Rattus norvegicus					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.					
AUTHORS	1 (bases 1 to 531)					
TITLE	Bonaldo,M.F., Lennon,G. and Soares,M.B.					
JOURNAL	Normalization and subtraction: two approaches to facilitate gene discovery					
MEDLINE	Genome Res. 6 (9), 791-806 (1996)					
COMMENT	97044477					
	Contact: Soares, MB					
	Program for Rat Gene Discovery and Mapping					
	University of Iowa					
	451 Eckstein Medical Research Building Iowa City, IA 52242, USA					
	Tel: 319 335 8250					
	Fax: 319 335 9565					
	Email: msoares@blue.weeg.uiowa.edu					
	The sequence contained an oligo-dT track that was present in the					
	oligonucleotide that was used to prime the synthesis of first					
	strand cDNA and therefore this may represent a bonafide poly A					





```

QY 62 TACCACAATATACAGGAGAATAAGTGAATGACAA 97
    ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 TCCAGAGTTCTAAAGGTGAATGGGTCTCATGGAA 351

RESULT 6
BI554667 776 bp mRNA linear EST 05-SEP-2001
LOCUS 603236619f1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5289605 5',
DEFINITION mRNA sequence.
ACCESSION BI554667
VERSION BI554667.1 GI:15441981
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 776)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11732 row: f column: 06
High quality sequence stop: 759.
FEATURES
    source
    1..776
    /organism="Mus musculus"
    /strain="129,C57BL/6J,FVB/N"
    /db_xref="taxon:10090"
    /clone="IMAGE:5289605"
    /clone_lib="NCI_CGAP_Mam3"
    /tissue_type="tumor, gross tissue"
    /dev_stage="10 months"
    /lab_host="DH10B"
    /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
    Site_2: NotI; Cloned unidirectionally. Primer: Oligo at.
    Library constructed by Life Technologies. Investigators
    providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
    Reference for transgenic model: Xu et al., Nature Genetics
    22, 37-43 (1999)."
BASE COUNT 175 a 233 c 231 g 137 t
ORIGIN

Query Match 35.4%; Score 36.8; DB 13; Length 776;
Best Local Similarity 61.5%; Pred. No. 6.7;
Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCTGTGCACCAAGTT 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 CAGGACAGTCAGGCTCCCTGGCTTCTCTGGACAGGAGGACCTGGAGTTCCAGGT 292

QY 62 TACCACAATATACAGGAGAATAAGTGAATGACAA 97
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 293 TCCAGAGTTCTAAAGGTGAATGGGTGTCATGGAA 328

RESULT 7
D69730 360 bp mRNA linear EST 07-DEC-1995
LOCUS CELK07307F Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone yk73c7 5', mRNA sequence.
ACCESSION D69730
VERSION D69730.1 GI:1105372
KEYWORDS EST.

```

```

SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 360)
AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
Tabara,H.
TITLE Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
FEATURES
    source
    1..360
    /organism="Caenorhabditis elegans"
    /strain="CB1489 hlm-8(e1489)"
    /db_xref="taxon:6239"
    /clone="yk73c7"
    /clone_lib="Yuji Kohara unpublished cDNA"
    /sex="hermaphrodite, male"
    /tissue_type="whole animal"
    /dev_stage="varied"
BASE COUNT 96 a 116 c 105 g 42 t 1 others
ORIGIN

Query Match 34.8%; Score 36.2; DB 14; Length 360;
Best Local Similarity 65.4%; Pred. No. 9.5;
Matches 53; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 1 GCTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCTGTGCACCAAGT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 177 GCTGGACCAACAGGGCCATCAGGAGCCCAAGGACCAAAAGGACCTTCAGGAGCCCCAGGA 236

QY 61 TTACCACAATATACAGGAGAA 81
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 237 GTCCAGGACAAACCGGAGGA 257

RESULT 8
BJI36441/c 632 bp mRNA linear EST 23-JAN-2002
LOCUS BJI36441 unpublished oligo-capped cDNA library, C. elegans L1 stage
DEFINITION Caenorhabditis elegans cDNA clone yk113b05 3', mRNA sequence.
ACCESSION BJI36441
VERSION BJI36441.1 GI:18296598
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 632)
AUTHORS Kohara,Y., Shin-i.T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE A complementary view of the C.elegans genome
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
FEATURES
    source
    1..632
    /organism="Caenorhabditis elegans"
    /strain="N2"
    /db_xref="taxon:6239"
    /clone="yk113b05"
    /clone_lib="unpublished oligo-capped cDNA library, C.

```

```

Db      548   GCTGGACCCAGGACCATTTCAGGAGCCCCCAGGACAAAAGGACTTCCAGGAGCCCAGGA    ||||||| ||||| | | | |||||| | | |||||| | | |||||| | | |||||| | | ||||||
QY      61     TTACCAACAATATACAGGAGA    81          ||| | | | ||||| I
Db      608   GCCCGAGCAATCCGGAGGA    628          ||| | | | ||||| I

RESULT 10
BI323061                                     461 bp        mRNA         linear       EST 30-JUN-2004
LOCUS   kt68h01.yl Strongyloides ratti L2 pAMP1 v1 Chiapelli McCarter
DEFINITION Strongyloides ratti cDNA 5' similar to TR:QI9079 QI9079 COSMID EGAP7. [1] ; contains element M8R1 repetitive element ; , mRNA sequence.
ACCESSION BI323061
VERSION   BI323061.1 GI:15002247
KEYWORDS Strongyloides ratti.
SOURCE    Strongyloides ratti.
ORGANISM  Strongyloides ratti.
REFERENCE
AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wyl, Wilson,R., Dante,M., Marra.M., Hillier,L., Kucaba,T., Theising,B., Bowe,Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarelshvili,I., Ronko,I., Kennedy,S., Maguire,L., Beck.C., Underwood,K., Steadman,M., Allen.M., Person.B., Swaller.T., Harvey.N., Schurk.R., Kalishin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Braxton,A.
TITLE     The Washington Univ. Nematode EST Project, 1999
JOURNAL   Unpublished (1999)
COMMENT   Contact: McCarter JP
           The Washington Univ. Nematode EST Project, 1999
           Washington University School of Medicine
           4444 Forest Park Parkway, Box 850L, St. Louis, MO 63108, USA
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: est@watson.wustl.edu
           The library was constructed by Brandi Chiapelli and Dr. James McCarter (chiapel@watson.wustl.edu & jmcarter@watson.wustl.edu) at Washington University, St. Louis. DNA Sequencing by: Washington University Genome Sequencing Center St. Louis.
           Seq primer: -4ORP from Gibco
           High quality sequence stop: 392.
FEATURES             Location/Qualifiers
     source            1.. .461
                     /organism="Strongyloides ratti"
                     /db_xref="taxon:34506"
                     /clone_lib="Strongyloides ratti L2 pAMP1 v1 Chiapellilibrary"
                     /McCarteer=""
                     /dev_stage="L2"
                     /lab_host="DH10B"
note=Vector: pAMP1 (Gibco); The library was constructed by Brandi Chiapelli and Dr. James McCarter at Washington University, St. Louis. The cDNA was made by using Dynabead oligo-dT priming (Dyna). PCR based library using a modified protocol from the SMART PCR CDNA Synthesis kit from clontech. Directionally cloned into UDG sites of pAMP1. Nematodes were provided by Dr. Mary Viney of Bristol, UK."

BASE COUNT   153 a      85 C      94 g      129 t
ORIGIN
```

Query Match                  34.6%; Score 36; DB 13; Length 461;  
Best Local Similarity    64.3%; Pred. No. 11;  
Matches         54; Conservative      0; Mismatches    30; Indels        0; Gaps

```

QY      4     GGACCTCCAGCACACGCCCGCCCGCCAGAGAGTAGTGCGGCCTCTGTGTCACCAGGTTTA    63
                                         ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| ||||| |||
Db      365   GGACCACCGAGGACCTCCAGGACCAACCAGGAAGATGGAAAGACGACCAGGACCATGATA    424A
                                         ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| ||||| |||
VY      64     CCACAATATACAGGAGAAAAAGT    87
                                         ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| ||||| |||
```

```

Db 425 CCAGGAATCCAGGACCACGAGT 448
|||| | || |||| | || ||
RESULT 11
LOCUS BG310518
DEFINITION HG310518 675 bp mRNA linear EST 23-FEB-2001
SWOV3MCAM56C07SK Onchocerca volvulus molting L3 larva cDNA
(SL96MLW-OvML3) Onchocerca volvulus cDNA clone SWOV3MCAM56C07 5',
mRNA sequence.
ACCESSION BG310518
VERSION BG310518.1 GI:13112376
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
AUTHORS Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL Unpublished (1997)
COMMENT Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genomesmith.edu
Seq primer: pluescript SK.
FEATURES
source
1..675
Location/Qualifiers
/organism="Onchocerca volvulus"
/strain="Kumba, Cameroon"
/db_xref="taxon:6282"
/clone="SWOV3MCAM56C07"
/cclone_lib="Onchocerca volvulus molting L3 larva cDNA
(SL96MLW-OvML3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/notes="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ML3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNA pol I. The library was constructed in the lambda
Uni-Zap XR vector and has 1 x 10E6 independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustigm@nyc.bc.org)."
BASE COUNT 190 a 157 c 166 g 162 t
ORIGIN
Query Match 34.4%; Score 35.8; DB 12; Length 675;
Best Local Similarity 63.2%; Pred. No. 12;
Matches 55; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
QY 1 GCTGGACCTCCAGCACACCCAGGCCCCAGAGAGTGGGGCCCTCTGGTCACGAGT 60
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 348 GTTGGACCACTGGACCACTGGCTCTCCAGGACGTGATGGACACCTGGACCAAGGA 407
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 TTACCAATAATACAGGAATAAAGT 87
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 408 CGACCAAGGTAAATCCAGATCACGAGT 434
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

---

```

RESULT 12
LOCUS BE029138
DEFINITION BE029138 287 bp mRNA linear EST 09-MAY-2001
kp25d01.y1.TBN95TM-SSFH Strongyloides stercoralis cDNA 5' similar
to WP:W01B6.7 CE03759 COL-2: CUTICLE COLLAGEN 2 ; contains element
MSRI repetitive element ;, mRNA sequence.
ACCESSION BE029138
VERSION BE029138.1 GI:8322600
KEYWORDS EST.
SOURCE Strongyloides stercoralis.
ORGANISM Strongyloides stercoralis.
REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Panagrolaimoidea; Strongyloidea; Strongyloidea.
AUTHORS McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.,
Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagaris,I., R.,
Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,
Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
Wilson,R.
TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Dr. Thomas Nutman and colleagues of
NIAID, NIH (tnutman@nih.gov). DNA Sequencing by: Washington
University Genome Sequencing Center St. Louis.
Seq primer: -40RP from Gibco.
FEATURES
source
1..287
Location/Qualifiers
/organism="Strongyloides stercoralis"
/strain="Filariform larvae obtained from humans"
/db_xref="taxon:6248"
/cclone_lib="TBN95TM-SSFH"
/lab_host="XLI-Blue MRF" (Stratagene)"
/notes="Vector: Lambda Uni-ZAP XR (Stratagene); Site_1:
EcoRI; Site_2: XhoI; mRNA was purified from 4 x 10E5
filariform larvae which had been isolated from infected
humans. cDNA was constructed and, using adaptors, was
cloned unidirectionally into the vector from the EcoRI
site to the XhoI site. The library has an unamplified
titer of 1.5 x 10E6 pfu/ml and an amplified, undiluted
titer of 7 x 10E9 pfu/ml. The average insert size of the
unamplified library is 975 bp (range, 500-1500)."
BASE COUNT 105 a 50 c 67 g 65 t
ORIGIN
Query Match 34.0%; Score 35.4; DB 10; Length 287;
Best Local Similarity 59.4%; Pred. No. 15;
Matches 60; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
QY 4 GGACCTCCAGCACACCCAGGCCCCAGAGAGTGGGGCCCTCTGGTCACGAGTTTA 63
||||| |||| | ||| |||| | ||| |||| | ||| |||| | ||| |||| |
Db 65 GGACCAACAGGACCACTGGACCAACAGGATCAGATGGAATCCAGAACTCCAGGACAA 124
|| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
QY 64 CCACAATATACAGGAGAAATAAGTGAAATGACAAATGCC 104
|| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 125 CCTGGACACAGGAGGCCAGGTGAAGGGAATCTGTCC 165
||||| |||| | |||| | |||| | |||| | |||| | |||| | |||| |

```

---

```

RESULT 13
LOCUS BE580165
DEFINITION BE580165 468 bp mRNA linear EST 09-MAY-2001
kq37h09.y1.TBN95TM-SSR Strongyloides stercoralis cDNA 5' similar to
TR:P79758 P79758 COLLAGEN ; contains TAR1.t3 TARI repetitive element
; , mRNA sequence.
ACCESSION BE580165

```

```

VERSON      BE580165.1  GI:98311107
SOURCE      EST.
ORGANISM    Strongyloides stercoralis.
REFERENCE   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
AUTHORS     Panagrolaimoidea; Strongyloidea; Strongyloides.
            1 (bases 1 to 468)
            McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,
            Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,
            Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagaris, V.,
            Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe
            M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,
            Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and
            Wilson, R.
            The Washington Univ. Nematode EST Project, 1999
            Unpublished (1999)
            Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            The library was constructed by Dr. Thomas Nutman and colleagues of
            NIAID, NIH (tnutman@nih.gov). DNA sequencing by: Washington
            University Genome Sequencing Center St. Louis.
            High quality sequence stop: 325.
            Location/Qualifiers
            1..468
            /organism="Strongyloides stercoralis"
            /strain="Rhabditiform larvae obtained from gerbils"
            /db_xref="taxon:6248"
            /clone_lib="TBN95TM-SSR"
            /lab_host="XL-1 Blue MRF" (Stratagene)
            /note="Vector: Lambda Uni-ZAP XR (Stratagene); Site_1:
            EcoRI; Site_2: XhoI; mRNA was purified from 2 x 10E3
            rhabditiform larvae which had been isolated from gerbils
            experimentally infected with larvae originally isolated
            from experimentally infected dogs. cDNA was constructed
            and, using adaptors, was cloned unidirectionally into the
            vector from the EcoRI site to the XhoI site. The library
            has an unamplified titer of 1 x 10E5 pfu/ml and an
            amplified, undiluted titer of 9 x 10E11 pfu/ml. The
            average insert size of the unamplified library is 675 bp
            (range, 100-1700)."
            BASE COUNT      157 a  92 c  134 g  85 t
            ORIGIN
            Query Match      33.8%; Score 35.2; DB 10; Length 468;
            Best Local Similarity 62.5%; Pred. NO. 18;
            Matches 55; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

            QY  4  GGACCTCCAGCACACCCAGGCCCCCAGAGAGAGTGGGGCTCTCGTGCACCAAGTTTA 63
            Db  157  GGACCAACAGGAGATCGATCTCCAGGAGAACAGGACCAAGGTGAACAGGTTTA 216
            QY  64  CCACATATACAGGAGAAATAGTGAAA 91
            Db  217  CAAGGAGATATTGGACCATCAGAGAAA 244

            RESULT 14
            LOCUS      BE579244
            DEFINITION  kq24e08.y1 TBN95TM-SSR Strongyloides stercoralis cDNA 5' similar to
            contains element PTR5 repetitive element ; mRNA sequence.
            ACCESSION  BE579244
            VERSION    BE579244.1  GI:9830186
            KEYWORDS   EST.
            SOURCE     Strongyloides stercoralis.
            ORGANISM   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Panagrolaimoidea; Strongyloidea; Strongyloides.

            TITLE      The Washington Univ. Nematode EST Project, 1999
            JOURNAL     Unpublished (1999)
            COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            The library was constructed by Dr. Thomas Nutman and colleagues of
            NIAID, NIH (tnutman@nih.gov). DNA sequencing by: Washington
            University Genome Sequencing Center St. Louis.
            High quality sequence stop: 325.
            Location/Qualifiers
            1..468
            /organism="Strongyloides stercoralis"
            /strain="Rhabditiform larvae obtained from gerbils"
            /db_xref="taxon:6248"
            /clone_lib="TBN95TM-SSR"
            /lab_host="XL-1 Blue MRF" (Stratagene)
            /note="Vector: Lambda Uni-ZAP XR (Stratagene); Site_1:
            EcoRI; Site_2: XhoI; mRNA was purified from 2 x 10E3
            rhabditiform larvae which had been isolated from gerbils
            experimentally infected with larvae originally isolated
            from experimentally infected dogs. cDNA was constructed
            and, using adaptors, was cloned unidirectionally into the
            vector from the EcoRI site to the XhoI site. The library
            has an unamplified titer of 1 x 10E5 pfu/ml and an
            amplified, undiluted titer of 9 x 10E11 pfu/ml. The
            average insert size of the unamplified library is 675 bp
            (range, 100-1700)."
            BASE COUNT      157 a  92 c  134 g  85 t
            ORIGIN
            Query Match      33.8%; Score 35.2; DB 10; Length 468;
            Best Local Similarity 62.5%; Pred. NO. 18;
            Matches 55; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

            QY  4  GGACCTCCAGCACACCCAGGCCCCCAGAGAGAGTGGGGCTCTCGTGCACCAAGTTTA 63
            Db  157  GGACCAACAGGAGATCGATCTCCAGGAGAACAGGACCAAGGTGAACAGGTTTA 216
            QY  64  CCACATATACAGGAGAAATAGTGAAA 91
            Db  217  CAAGGAGATATTGGACCATCAGAGAAA 244

            RESULT 15
            LOCUS      BE286914
            DEFINITION  BE286914.1  GI:9165563
            ACCESSION  BE286914
            VERSION    BE286914.1  GI:9165563
            KEYWORDS   EST.
            SOURCE     house mouse.
            ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            NIH-MGC http://mgc.nci.nih.gov/.
            1 (bases 1 to 814)
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs@mail.nih.gov
            TITLE      The Washington Univ. Nematode EST Project, 1999
            JOURNAL     Unpublished (1999)
            COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            The library was constructed by Dr. Thomas Nutman and colleagues of
            NIAID, NIH (tnutman@nih.gov). DNA sequencing by: Washington
            University Genome Sequencing Center St. Louis.
            High quality sequence stop: 432.
            Location/Qualifiers
            1..483
            /organism="Strongyloides stercoralis"
            /strain="Rhabditiform larvae obtained from gerbils"
            /db_xref="taxon:6248"
            /clone_lib="TBN95TM-SSR"
            /lab_host="XL-1 Blue MRF" (Stratagene)
            /note="Vector: Lambda Uni-ZAP XR (Stratagene); Site_1:
            EcoRI; Site_2: XhoI; mRNA was purified from 2 x 10E3
            rhabditiform larvae which had been isolated from gerbils
            experimentally infected with larvae originally isolated
            from experimentally infected dogs. cDNA was constructed
            and, using adaptors, was cloned unidirectionally into the
            vector from the EcoRI site to the XhoI site. The library
            has an unamplified titer of 1 x 10E5 pfu/ml and an
            amplified, undiluted titer of 9 x 10E11 pfu/ml. The
            average insert size of the unamplified library is 675 bp
            (range, 100-1700)."
            BASE COUNT      175 a  88 c  131 g  89 t
            ORIGIN
            Query Match      33.8%; Score 35.2; DB 10; Length 483;
            Best Local Similarity 62.5%; Pred. NO. 18;
            Matches 55; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

            QY  4  GGACCTCCAGCACACCCAGGCCCCCAGAGAGAGTGGGGCTCTCGTGCACCAAGTTTA 63
            Db  23  GGACCAACAGGAGATCGATCTCCAGGAGAACAGGACCAAGGTGAACAGGTTTA 82
            QY  64  CCACATATACAGGAGAAATAGTGAAA 91
            Db  83  CAAGGAGATATTGGACCATCAGAGAAA 110

            RESULT 15
            LOCUS      BE286914
            DEFINITION  BE286914.1  GI:9165563
            ACCESSION  BE286914
            VERSION    BE286914.1  GI:9165563
            KEYWORDS   EST.
            SOURCE     house mouse.
            ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            NIH-MGC http://mgc.nci.nih.gov/.
            1 (bases 1 to 814)
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs@mail.nih.gov
```

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM8524 row: h column: 19  
High quality sequence start: 9  
High quality sequence stop: 604.  
Location/Qualifiers  
1. .814  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3487170"  
/clone\_lib="NCI\_CGAP\_Mam5"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="7 months"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigators  
providing samples: Lothar Hennighausen/Robin Humphreys,  
NIH"

BASE COUNT	213 a	213 c	279 g	108 t	1 others
ORIGIN					
Query Match	33.8%; Score 35.2; DB 10; Length 814;				
Best Local Similarity	60.4%; Pred. No. 18;				
Matches	58; Conservative	0; Mismatches	38; Indels	0; Gaps	0;
QY	2	CTGGACCTCCAGCACACCCCGCCGAGAGAAAGTGGGGCCTCCTGTGGTGCACCGGTT	61		
Db	166	CAGGACAGTCAGGCCCTCCCTGGCCTTCCTGGACAGAGGGGACACCTGGAGTCCAGGTT	225		
QY	62	TACCACAATATACAGGAGAAATAAGTGAATGACAA	97		
Db	226	TCCCAGGTTCTAAAGGTGAATGGGCGTCATGGGAA	261		

Search completed: February 20, 2003, 06:14:21  
Job time : 264.925 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 16:09:30 ; Search time 6.86562 Seconds  
(without alignments)  
4645.518 Million cell updates/sec

Title: us-09-997-610-1\_COPY\_53\_156

Perfect score: 104

Sequence: 1 gctggacctccagcacccc.....agtgaataatgacaaaatgcc 104

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*

4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*

5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33.2	31.9	1881	4	US-09-029-348-20
2	30.4	29.2	821	4	US-09-342-681C-7
3	30.4	29.2	1176	4	US-09-342-681C-14
4	30.4	29.2	1574	4	US-09-342-681C-1
5	30.4	29.2	3394	1	US-08-159-784-4
6	28.8	27.7	1898	1	US-08-342-411A-1
7	28.4	27.3	1839	1	US-08-383-744-1
8	28.4	27.3	1839	2	US-08-999-336-1
9	28.4	27.3	1839	5	PCT-US96-01427-1
c 10	28.2	27.1	390	4	US-09-134-001C-635
c 11	28.2	27.1	513	4	US-09-134-001C-647
12	28.2	27.1	585	4	US-09-134-001C-726
c 13	28.2	27.1	810	4	US-09-134-001C-624
14	28.2	27.1	3552	4	US-09-134-001C-693
15	28	26.9	1588	6	5510466-3
16	28	26.9	1507	6	5510466-1
17	28	26.9	5102	6	5510466-1
18	27.6	26.5	1074	2	US-08-494-168-1
19	27.6	26.5	1404	6	US-08-627-151A-15
20	27.6	26.5	1404	6	5480796-8
21	27.6	26.5	1486	4	US-08-795-473B-3
22	27.6	26.5	1486	4	US-09-439-856-3
23	27.6	26.5	2061	6	5171840-1
24	27.6	26.5	2061	6	5480796-1
25	27.6	26.5	3319	4	US-08-795-473B-2
26	27.6	26.5	3319	4	US-09-439-856-2
27	27.4	26.3	2543	1	US-08-555-669-11

ALIGNMENTS

RESULT 1

US-09-029-348-20

; Sequence 20, Application US/09029348

; Patent No. 6171827

; GENERAL INFORMATION:

; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER

; TITLE OF INVENTION: NOVEL PROCOLLAGENS

; FILE REFERENCE: G087857PUS LISTING

; CURRENT APPLICATION NUMBER: US/09/029,348

; CURRENT FILING DATE: 1998-05-07

; NUMBER OF SEQ ID NOS: 20

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 20

; LENGTH: 1881

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE

; OTHER INFORMATION: DERIVED FROM cDNA OF PROCOLLAGENS

US-09-029-348-20

Query Match

Best Local Similarity 31.9%; Score 33.2; DB 4; Length 1881;

Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 2 CTGACCTCCAGCACACCCAGGCCCCAGAGAGAGTGGGCGCTCTGGTGACACAGGTT 61

Db 581 CTGGATCTCCAGGATACCAAGGACCCCTGGTGAACCTGGCAAGCTGGCTTCAGGCC 640

Qy 62 TACCACATATACAGGAGAAATAGT 87

Db 641 CTCAGGACCTCTGGTGCTATAGT 666

RESULT 2

US-09-342-681C-7

; Sequence 7, Application US/09342681C

; Patent No. 6355782

; GENERAL INFORMATION:

; APPLICANT: Zonana et al.

; TITLE OF INVENTION: Hypohydrotic ectodermal dysplasia genes and proteins

; FILE REFERENCE: 52978

; CURRENT APPLICATION NUMBER: US/09/342,681C

; CURRENT FILING DATE: 1999-06-29

; PRIOR APPLICATION NUMBER: 60/092,279

; PRIOR FILING DATE: 1998-07-09

; PRIOR APPLICATION NUMBER: 60/112,366

; PRIOR FILING DATE: 1998-12-15

; NUMBER OF SEQ ID NOS: 123

; SOFTWARE: PatentIn Ver. 2.1



RESULT 6  
US-08-342-411A-1  
; Sequence 1, Application US/08342411A  
; Patent No. 5639616  
; GENERAL INFORMATION:  
; APPLICANT: LIAO, Shutsung  
; APPLICANT: SONG, Ching  
; TITLE OF INVENTION: UBIQUITOUS NUCLEAR RECEPTOR:  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS  
; NUMBER OF SEQUENCES: 38  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/342.411A  
; FILING DATE: 18-NOV-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KITCHELL, BARBARA S.  
; REGISTRATION NUMBER: 33,928  
; REFERENCE/DOCKET NUMBER: ARCD154  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (713) 789-2679  
; TELEX: 79-0924  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1898 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 71..1450  
US-08-342-411A-1

Query Match 27.7%; Score 28.8; DB 1; Length 1898;  
Best Local Similarity 56.2%; Pred. No. 2.4;  
Matches 54; Conservative 0; Mismatches 42; Indels 0; Gaps 0;  
Qy 4 GGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCTGTGCACAGGTTTA 63  
Db 1749 GGACCATGGCTCTCTCCCTAGCCCGGAGACAGGCGCTCTCTCTCTCTGT 1808  
Qy 64 CCACATATACAGGAGAATAAGTGAATGACAAA 99  
Db 1809 ATTTAATAAACTAAACACAGAAAAA 1844

RESULT 7  
US-08-383-744-1  
; Sequence 1, Application US/08383744  
; Patent No. 5702948  
; GENERAL INFORMATION:  
; APPLICANT: Greene, Mark I.  
; APPLICANT: Davis, James G.  
; TITLE OF INVENTION: Saccular collagen and Compositions  
; TITLE OF INVENTION: and  
; TITLE OF INVENTION: Methods for Making and Using the Same  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &  
; STREET: One Liberty Place, 46th floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/999,336  
; FILING DATE:  
; CLASSIFICATION:

; STREET: One Liberty Place, 46th floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/383,744  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: DeLuca, Mark  
; REGISTRATION NUMBER: 33,229  
; REFERENCE/DOCKET NUMBER: UPN-2039  
; TELEPHONE: 215-568-3100  
; TELEFAX: 215-568-3439  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1839 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: both  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 331..1602  
US-08-383-744-1  
Query Match 27.3%; Score 28.4; DB 1; Length 1839;  
Best Local Similarity 66.1%; Pred. No. 3.2;  
Matches 41; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
Qy 28 CCAGAAGAAGTGGGGCTCTGTGCACAGGTTTACCACAAATATACAGAGAAATAAGT 87  
Db 670 CAAGGACCGGTGGTCTCTCTGTGTCCTCCAGGAAACCCAGGAGAGAAAGT 729  
Qy 88 GA 89  
Db 730 GA 731  
RESULT 8  
US-08-999-336-1  
; Sequence 1, Application US/08999336  
; Patent No. 5891850  
; GENERAL INFORMATION:  
; APPLICANT: Greene, Mark I.  
; APPLICANT: Davis, James G.  
; TITLE OF INVENTION: Saccular collagen and Compositions  
; TITLE OF INVENTION: and Methods for Making and Using the Same  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &  
; STREET: One Liberty Place, 46th floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/999,336  
; FILING DATE:  
; CLASSIFICATION:

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/383,744
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: UPN-2039
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1839 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 331..1602
; PCT-US96-01427-1
```

```
Query Match 27.3%; Score 28.4; DB 2; Length 1839;
Best Local Similarity 66.1%; Pred. No. 3.2;
Matches 41; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
```

```
QY 28 CCAGAAGAAGTGGGCTCTCTGTGCACCGAGTTTACCACAATATACAGGAGAAATAAGT 87
Db 670 CAAGGACCGGTGGTCTTCTGTGTCTCCCTGGATCCCGAGAAACACGAGAGAAAGGT 729

QY 88 GA 89
Db 730 GA 731
```

```
RESULT 9
PCT-US96-01427-1
; Sequence 1, Application PC/TUS9601427
; GENERAL INFORMATION:
; APPLICANT: Greene, Mark I.
; APPLICANT: Davis, James G.
; TITLE OF INVENTION: Saccular collagen and Compositions and
; TITLE OF INVENTION: Methods for Making and Using the Same
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/01427
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/383,744
; FILING DATE: 02-FEB-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: UPN-2653
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 1839 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 331..1602
; PCT-US96-01427-1
```

```
Query Match 27.3%; Score 28.4; DB 5; Length 1839;
Best Local Similarity 66.1%; Pred. No. 3.2;
Matches 41; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
```

```
QY 28 CCAGAAGAAGTGGGCTCTCTGTGCACCGAGTTTACCACAATATACAGGAGAAATAAGT 87
Db 670 CAAGGACCGGTGGTCTTCTGTGTCTCCCTGGATCCCGAGAAACACGAGAGAAAGGT 729

QY 88 GA 89
Db 730 GA 731
```

```
RESULT 10
US-09-134-001C-635/C
; Sequence 635, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 635
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-635
```

```
Query Match 27.1%; Score 28.2; DB 4; Length 390;
Best Local Similarity 57.3%; Pred. No. 2.2;
Matches 51; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
```

```
QY 2 CTGGACCTCCAGCACACCCCGCCAGCCCGCCAGAAAGTGGGGCTCTCTGTGTGCACAGGTT 61
Db 161 CAGGTACGCCAGCAGCAACCCAGGTACGCCAGCAGCAACCCAGGTAACCCAGCGGAACAGGTA 102

QY 62 TACCACAATATACAGGAGAAATAAGTGAA 90
Db 101 CGCCAGCAGAACCCAGGTAAACCCAGCGGAA 73
```

```
RESULT 11
US-09-134-001C-647/c
; Sequence 647, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
```



	Query Match	26.9%	Score 28	DB 6	Length 1588
	Best Local Similarity	58.3%	Pred. No. 4.1		
	Matches 49	Conservative 0	Mismatches 35	Indels 0	Gaps 0
Qy	4	GGACCTCCAGCACACCCCGCCCGCCAGACAGAACTGGGGCCTCTCTGGTGCACACAGGTTTA	63		
Db	817	GGTCCTCTGGACCTTCACAGGTGAAAAAGGAGATAGAGGCCCTCTCGGACAAAATGGTATA	876		
Qy	64	CCACAATATACAGGAGAAATAAGT	87		
Db	877	CCAGGCTTTCAGGCTCTAATAGGT	900		

Search completed: February 19, 2003, 22:58:49  
Job time : 10.8656 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 18:01:21 ; Search time 8.36875 Seconds  
(without alignments)  
6329.459 Million cell updates/sec

Title: US-09-997-610-1\_COPY\_53\_156

Perfect score: 104

Sequence: 1 gctgacactccagcacaccc.....agtgaatgacaaaatgcc 104

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 424239 seqs, 254661826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*

2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*

3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*

4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*

5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*

6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*

7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*

8: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*

9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*

10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*

11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*

12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*

13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*

14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	104	100.0	1381	9	US-09-997-610-1
2	104	100.0	1731	9	US-09-997-610-5
3	75	72.1	1377	9	US-09-997-610-3
4	75	72.1	1731	9	US-09-997-610-7
5	33.2	31.9	561	10	US-09-923-779-47
6	33.2	31.9	617	10	US-09-923-779-51
7	33.2	31.9	648	10	US-09-923-779-66
8	33.2	31.9	653	10	US-09-923-779-53
9	33.2	31.9	655	10	US-09-923-779-68
10	33.2	31.9	656	10	US-09-923-779-69
11	33.2	31.9	658	10	US-09-923-779-71
12	33.2	31.9	659	10	US-09-923-779-52
13	33.2	31.9	780	10	US-09-923-779-5
14	33.2	31.9	823	10	US-09-923-779-28
15	32.8	31.5	392	10	US-09-960-352-1786
16	32.8	31.5	415	10	US-09-960-352-9137
17	32.8	31.5	549	10	US-09-923-779-49
18	32.8	31.5	6158	10	US-09-919-497-6
19	32.8	31.5	6158	10	US-09-954-456-762

20	32.4	31.2	886	10	US-09-923-779-4	Sequence 4, Appli
21	31.2	30.0	1619	9	US-09-764-868-400	Sequence 400, App
22	31	29.8	3226	10	US-09-954-456-725	Sequence 725, App
23	30.4	29.2	821	9	US-09-729-6588-7	Sequence 7, Appli
24	30.4	29.2	1176	9	US-09-729-6588-14	Sequence 14, Appli
25	30.4	29.2	1574	9	US-09-729-6588-1	Sequence 1, Appli
26	30.4	29.2	2866	9	US-09-764-868-48	Sequence 48, Appli
27	30.4	29.2	3380	10	US-09-799-799-1	Sequence 1, Appli
28	30.4	29.2	3394	10	US-09-880-107-2178	Sequence 2178, Ap
29	30.2	29.0	88191	10	US-09-799-799-3	Sequence 3, Appli
30	29.4	28.3	1320	10	US-09-815-242-7615	Sequence 7615, Ap
31	28.8	27.7	1155	10	US-09-833-381-1293	Sequence 1293, Ap
32	28.6	27.5	684973	10	US-09-263-959-1	Sequence 1, Appli
33	28.4	27.3	572	10	US-09-923-779-56	Sequence 56, Appli
34	28	26.9	1234	9	US-09-954-531-1366	Sequence 44, Appli
35	28	26.9	1344	10	US-09-925-299-44	Sequence 25, Appli
36	28	26.9	1969	10	US-09-789-561-25	Sequence 49, Appli
37	27.8	26.7	259	10	US-09-765-231A-49	Sequence 25, Appli
38	27.6	26.5	3477	9	US-09-935-868-25	Sequence 23, Appli
39	27.6	26.5	3507	9	US-09-935-868-23	Sequence 6, Appli
40	27.6	26.5	6158	10	US-09-919-497-6	Sequence 762, App
41	27.6	26.5	6158	10	US-09-954-456-762	Sequence 1494, Ap
42	27.6	26.5	32183	10	US-09-764-869-1494	Sequence 7, Appli
43	27.4	26.3	2823	10	US-09-919-497-7	Sequence 448, App
44	27.4	26.3	3690	12	US-10-044-090-448	Sequence 4300, Ap
45	27.2	26.2	274	10	US-09-923-876-4300	

#### ALIGNMENTS

##### RESULT 1

US-09-997-610-1

; Sequence 1, Application US/09997610

; Patent No. US20020156244A1

; GENERAL INFORMATION:

; APPLICANT: Fox, Brian

; APPLICANT: Holloway, James L.

; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN

; FILE REFERENCE: ZACRR13

; CURRENT APPLICATION NUMBER: US/09/997,610

; PRIOR FILING DATE: 2001-11-29

; PRIOR APPLICATION NUMBER: US 60/253,924

; PRIOR FILING DATE: 2000-11-29

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 1381

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (2)...(1381)

US-09-997-610-1

Query Match 100.0%; Score 104; DB 9; Length 1381;  
Best Local Similarity 100.0%; Pred. No. 1.4e-24;  
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCTGACCTCCAGCACACCCAGGGCCCCAGAGAGTGGGGCCCTCTGGTGCACACAGGT 60

|||||

Db 53 GCTGACCTCCAGCACACCCAGGGCCCCAGAGAGTGGGGCCCTCTGGTGCACACAGGT 112

|||||

QY 61 TTACCAATATACAGGAGAAATAGTGAATGACAAAATGCC 104

|||||

Db 113 TTACCAATATACAGGAGAAATAGTGAATGACAAAATGCC 156

|||||

RESULT 2

US-09-997-610-5

; Sequence 5, Application US/09997610

; Patent No. US20020156244A1

GENERAL INFORMATION:  
; APPLICANT: Fox, Brian  
; APPLICANT: Holloway, James L.  
; TITLE OF INVENTION: ADIPOCYTE COMPLEMENT RELATED PROTEIN  
; TITLE OF INVENTION: ZACRP13  
; FILE REFERENCE: 00-96  
; CURRENT APPLICATION NUMBER: US/09/997,610  
; PRIOR FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: US 60/253,924  
; PRIOR FILING DATE: 2000-11-29  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 1731  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1731)  
US-09-997-610-5

Query Match 100.0%; Score 104; DB 9; Length 1731;  
Best Local Similarity 100.0%; Pred. No. 1.5e-24;  
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCTGGACCTCCAGCACACCCAGGCCGCCAGAGAAGTGGGGCTCTCTGTGCACCAAGT 60  
DB 406 GCTGGACCTCCAGCACACCCAGGCCGCCAGAGAAGTGGGGCTCTCTGTGCACCAAGT 465  
QY 61 TTACCACAATATACAGGAGAAATAAGTGAATGACAAAATGCC 104  
DB 466 TTACCACAATATACAGGAGAAATAAGTGAATGACAAAATGCC 509

RESULT 3  
US-09-997-610-3  
; Sequence 3, Application US/09997610  
; Patent No. US20020156244A1  
; GENERAL INFORMATION:  
; APPLICANT: Fox, Brian  
; APPLICANT: Holloway, James L.  
; TITLE OF INVENTION: ADIPOCYTE COMPLEMENT RELATED PROTEIN  
; TITLE OF INVENTION: ZACRP13  
; FILE REFERENCE: 00-96  
; CURRENT APPLICATION NUMBER: US/09/997,610  
; CURRENT FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: US 60/253,924  
; PRIOR FILING DATE: 2000-11-29  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 1377  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Degenerate polynucleotide encoding a polypeptide  
; OTHER INFORMATION: of SEQ ID NO:2  
; NAME/KEY: variation  
; LOCATION: (1)...(1377)  
; OTHER INFORMATION: Each n is independently A, T, G, or C.  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1377)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-997-610-3

Query Match 72.1%; Score 75; DB 9; Length 1377;  
Best Local Similarity 64.4%; Pred. No. 3.8e-15;  
Matches 67; Conservative 14; Mismatches 23; Indels 0; Gaps 0;  
QY 1 GCTGGACCTCCAGCACACCCAGGCCGCCAGAGAAGTGGGGCTCTCTGTGCACCAAGT 60  
DB 52 GCNGGNCNCNCNCAYCCNMCCNGCARGCTNGCNCNGCNCNGCNCNGN 111

QY 61 TTACCACAATATACAGGAGAAATAAGTGAATGACAAAATGCC 104  
DB 112 YTNCCNCARTAYACNGGNGARATHWSNGARATGACNAARTGYCC 155  
RESULT 4  
US-09-997-610-7  
; Sequence 7, Application US/09997610  
; Patent No. US20020156244A1  
; GENERAL INFORMATION:  
; APPLICANT: Fox, Brian  
; APPLICANT: Holloway, James L.  
; TITLE OF INVENTION: ADIPOCYTE COMPLEMENT RELATED PROTEIN  
; TITLE OF INVENTION: ZACRP13  
; FILE REFERENCE: 00-96  
; CURRENT APPLICATION NUMBER: US/09/997,610  
; CURRENT FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: US 60/253,924  
; PRIOR FILING DATE: 2000-11-29  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 7  
; LENGTH: 1731  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Degenerate polynucleotide sequence of zacrp13/zhp1  
; OTHER INFORMATION: of SEQ ID NO:6  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1731)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-997-610-7

Query Match 72.1%; Score 75; DB 9; Length 1731;  
Best Local Similarity 64.4%; Pred. No. 4.2e-15;  
Matches 67; Conservative 14; Mismatches 23; Indels 0; Gaps 0;  
QY 1 GCTGGACCTCCAGCACACCCAGGCCGCCAGAGAAGTGGGGCTCTCTGTGCACCAAGT 60  
DB 406 GCNGGNCNCNCNCAYCCNMCCNGCARGCTNGCNCNGCNCNGCNCNGN 465  
QY 61 TTACCACAATATACAGGAGAAATAAGTGAATGACAAAATGCC 104  
DB 466 YTNCCNCARTAYACNGGNGARATHWSNGARATGACNAARTGYCC 509

RESULT 5  
US-09-923-779-47  
; Sequence 47, Application US/09923779  
; Patent No. US20020076721A1  
; GENERAL INFORMATION:  
; APPLICANT: Pyle, Ruth A.  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER  
; FILE REFERENCE: 210121.553  
; CURRENT APPLICATION NUMBER: US/09/923,779  
; CURRENT FILING DATE: 2001-08-06  
; NUMBER OF SEQ ID NOS: 155  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 47  
; LENGTH: 561  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 544, 550  
; OTHER INFORMATION: n = A,T,C or G  
US-09-923-779-47

Query Match 31.9%; Score 33.2; DB 10; Length 561;  
Best Local Similarity 61.6%; Pred. No. 0.11;

```
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCTGGTGACCAAGTT 61
||||| ||||| ||| ||||| ||| ||||| ||||| |||||
Db 404 CTGGATCTCCAGGATACCAAGGACCCCTGGTGAACCTGGGCAAGCTGGTCTTCAGGCC 463
QY 62 TACCACAATATACAGGAGAAATAAGT 87
||||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464 CTCCAGGACCTCTCTGGTGTATAGT 489

RESULT 6
US-09-923-779-51
; Sequence 51, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51
; LENGTH: 617
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 581, 605
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-51

Query Match 31.9%; Score 33.2; DB 10; Length 617;
Best Local Similarity 61.6%; Pred. No. 0.12;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCTGGTGACCAAGTT 61
||||| ||||| ||| ||||| ||| ||||| ||||| |||||
Db 404 CTGGATCTCCAGGATACCAAGGACCCCTGGTGAACCTGGGCAAGCTGGTCTTCAGGCC 463
QY 62 TACCACAATATACAGGAGAAATAAGT 87
||||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464 CTCCAGGACCTCTCTGGTGTATAGT 489

RESULT 7
US-09-923-779-66
; Sequence 66, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 648
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 642, 646
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-66

Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCTGGTGACCAAGTT 61
||||| ||||| ||| ||||| ||| ||||| ||||| |||||
Db 404 CTGGATCTCCAGGATACCAAGGACCCCTGGTGAACCTGGGCAAGCTGGTCTTCAGGCC 463
QY 62 TACCACAATATACAGGAGAAATAAGT 87
||||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464 CTCCAGGACCTCTCTGGTGTATAGT 489

RESULT 8
US-09-923-779-53
; Sequence 53, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 53
; LENGTH: 653
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 517, 579, 581, 603, 649
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-53

Query Match 31.9%; Score 33.2; DB 10; Length 653;
Best Local Similarity 61.6%; Pred. No. 0.12;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCTGGTGACCAAGTT 61
||||| ||||| ||| ||||| ||| ||||| ||||| |||||
Db 404 CTGGATCTCCAGGATACCAAGGACCCCTGGTGAACCTGGGCAAGCTGGTCTTCAGGCC 463
QY 62 TACCACAATATACAGGAGAAATAAGT 87
||||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464 CTCCAGGACCTCTCTGGTGTATAGT 489

RESULT 9
US-09-923-779-68
; Sequence 68, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 655
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
```

```
Query Match 31.9%; Score 33.2; DB 10; Length 648;
Best Local Similarity 61.6%; Pred. No. 0.12;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCTGGTGACCAAGTT 61
||||| ||||| ||| ||||| ||| ||||| ||||| |||||
Db 404 CTGGATCTCCAGGATACCAAGGACCCCTGGTGAACCTGGGCAAGCTGGTCTTCAGGCC 463
QY 62 TACCACAATATACAGGAGAAATAAGT 87
||||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464 CTCCAGGACCTCTCTGGTGTATAGT 489

RESULT 8
US-09-923-779-53
; Sequence 53, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 53
; LENGTH: 653
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 517, 579, 581, 603, 649
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-53

Query Match 31.9%; Score 33.2; DB 10; Length 653;
Best Local Similarity 61.6%; Pred. No. 0.12;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCTGGTGACCAAGTT 61
||||| ||||| ||| ||||| ||| ||||| ||||| |||||
Db 404 CTGGATCTCCAGGATACCAAGGACCCCTGGTGAACCTGGGCAAGCTGGTCTTCAGGCC 463
QY 62 TACCACAATATACAGGAGAAATAAGT 87
||||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464 CTCCAGGACCTCTCTGGTGTATAGT 489

RESULT 9
US-09-923-779-68
; Sequence 68, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 655
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
```

; LOCATION: 654  
; OTHER INFORMATION: n = A,T,C or G  
US-09-923-779-68

Query Match 31.9%; Score 33.2; DB 10; Length 655;  
Best Local Similarity 61.6%; Pred. No. 0.12; Indels 0; Gaps 0;  
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
QY 2 CTGGACCTCCAGCACACCCAGGCCGCCAGAGAAGTGGGCTCTCTGTGCACCAAGTT 61  
DB 404 CTGGATCTCCAGATACCAAGACCCCTGGTGAACCTGGGCAAGCTGCTCTCAGGCC 463  
QY 62 TACCACAATATACAGAGAAATAAGT 87  
DB 464 CTCAGGACCTCTGTGCTATAGT 489

RESULT 10  
US-09-923-779-69  
; Sequence 69, Application US/09923779  
; Patent No. US20020076721A1  
; GENERAL INFORMATION:  
; APPLICANT: Pyle, Ruth A.  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.553  
; CURRENT APPLICATION NUMBER: US/09/923,779  
; CURRENT FILING DATE: 2001-08-06  
; NUMBER OF SEQ ID NOS: 155  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 69  
; LENGTH: 656  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 565, 619, 621  
; OTHER INFORMATION: n = A,T,C or G  
US-09-923-779-69

Query Match 31.9%; Score 33.2; DB 10; Length 656;  
Best Local Similarity 61.6%; Pred. No. 0.12; Indels 0; Gaps 0;  
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
QY 2 CTGGACCTCCAGCACACCCAGGCCGCCAGAGAAGTGGGCTCTCTGTGCACCAAGTT 61  
DB 404 CTGGATCTCCAGATACCAAGACCCCTGGTGAACCTGGGCAAGCTGCTCTCAGGCC 463  
QY 62 TACCACAATATACAGAGAAATAAGT 87  
DB 464 CTCAGGACCTCTGTGCTATAGT 489

RESULT 11  
US-09-923-779-71  
; Sequence 71, Application US/09923779  
; Patent No. US20020076721A1  
; GENERAL INFORMATION:  
; APPLICANT: Pyle, Ruth A.  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.553  
; CURRENT APPLICATION NUMBER: US/09/923,779  
; CURRENT FILING DATE: 2001-08-06  
; NUMBER OF SEQ ID NOS: 155  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 71  
; LENGTH: 658  
; TYPE: DNA

; ORGANISM: Homo sapiens  
US-09-923-779-71

Query Match 31.9%; Score 33.2; DB 10; Length 658;  
Best Local Similarity 61.6%; Pred. No. 0.12; Indels 0; Gaps 0;  
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
QY 2 CTGGACCTCCAGCACACCCAGGCCGCCAGAGAAGTGGGCTCTCTGTGCACCAAGTT 61  
DB 404 CTGGATCTCCAGATACCAAGACCCCTGGTGAACCTGGGCAAGCTGCTCTCAGGCC 463  
QY 62 TACCACAATATACAGAGAAATAAGT 87  
DB 464 CTCAGGACCTCTGTGCTATAGT 489

RESULT 12  
US-09-923-779-52  
; Sequence 52, Application US/09923779  
; Patent No. US20020076721A1  
; GENERAL INFORMATION:  
; APPLICANT: Pyle, Ruth A.  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.553  
; CURRENT APPLICATION NUMBER: US/09/923,779  
; CURRENT FILING DATE: 2001-08-06  
; NUMBER OF SEQ ID NOS: 155  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 52  
; LENGTH: 659  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 627  
; OTHER INFORMATION: n = A,T,C or G  
US-09-923-779-52

Query Match 31.9%; Score 33.2; DB 10; Length 659;  
Best Local Similarity 61.6%; Pred. No. 0.12; Indels 0; Gaps 0;  
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
QY 2 CTGGACCTCCAGCACACCCAGGCCGCCAGAGAAGTGGGCTCTCTGTGCACCAAGTT 61  
DB 404 CTGGATCTCCAGATACCAAGACCCCTGGTGAACCTGGGCAAGCTGCTCTCAGGCC 463  
QY 62 TACCACAATATACAGAGAAATAAGT 87  
DB 464 CTCAGGACCTCTGTGCTATAGT 489

RESULT 13  
US-09-923-779-5/c  
; Sequence 5, Application US/09923779  
; Patent No. US20020076721A1  
; GENERAL INFORMATION:  
; APPLICANT: Pyle, Ruth A.  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.553  
; CURRENT APPLICATION NUMBER: US/09/923,779  
; CURRENT FILING DATE: 2001-08-06  
; NUMBER OF SEQ ID NOS: 155  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 780  
; TYPE: DNA  
; ORGANISM: Homo sapiens

```

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 456, 489, 532, 562, 608, 622, 628, 640, 659, 672, 675, 701,
; LOCATION: 704, 712, 718, 772, 779
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-5

Query Match      31.9%; Score 33.2; DB 10; Length 780;
Best Local Similarity 61.6%; Pred. No. 0.13;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy  2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAGAGTGGGGCTCTCTGGTGCCACCAAGTTT 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  392 CTGGATCTCCAGGATACCAAGGACCCCTCTGTGAACCTGGGCAAGCTGGTCTTCAGGCC 333
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  62 TACCACAATATACAGGAGAAATAAGT 87
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  332 CTCAGGACCTCCCTGGTGCTATAGT 307
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
US-09-923-779-28
; Sequence 28 Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923.779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 823
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-923-779-28

Query Match      31.9%; Score 33.2; DB 10; Length 823;
Best Local Similarity 61.6%; Pred. No. 0.13;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy  2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAGAGTGGGGCTCTCTGGTGCCACCAAGTTT 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  413 CTGGATCTCCAGGATACCAAGGACCCCTCTGTGAACCTGGGCAAGCTGGTCTTCAGGCC 472
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  62 TACCACAATATACAGGAGAAATAAGT 87
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  473 CTCAGGACCTCTCTGGTGCTATAGT 498
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
US-09-960-352-1786
; Sequence 1786 Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960.352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 1786
; LENGTH: 392
; TYPE: DNA
; ORGANISM: Bos taurus
```

```

; OTHER INFORMATION: Clone ID: 08-LIB34-023-Q1-E1-B7
US-09-960-352-1786

Query Match      31.5%; Score 32.8; DB 10; Length 392;
Best Local Similarity 64.5%; Pred. No. 0.13;
Matches 49; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy  2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAGAGTGGGGCTCTCTGGTGCCACCAAGTTT 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  153 CGGGGCCACCAAGGACACCCAGGCCCTCCAGGTATTAAGAGGGCTCCAGGTATAACAGGAA 212
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  62 TACCACAATATACAGG 77
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  213 TACCAGGTTTGCAGG 228
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: February 19, 2003, 23:02:31
Job time : 11.3688 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 15:08:07 ; Search time 35.9937 Seconds  
(without alignments)  
6506.903 Million cell updates/sec

Title: us-09-997-610-1-copy\_53\_156

Perfect score: 104

Sequence: 1 gctggacctccagcacccc.....agtgaataacacaaatgccc 104

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59	56.7	1338	24	Human genset metab
2	36.8	35.4	6512	24	Mouse ischaemic co
3	34	32.7	5467	22	Human EST-derived
4	34	32.7	5468	22	Human EST-derived
5	33.2	31.9	561	24	CDNA #47 encoding
6	33.2	31.9	617	24	CDNA #51 encoding
7	33.2	31.9	648	24	CDNA #66 encoding
8	33.2	31.9	653	24	CDNA #53 encoding
9	33.2	31.9	655	24	CDNA #68 encoding

10	33.2	31.9	656	24	ABK44129	CDNA #69 encoding
11	33.2	31.9	658	24	ABK44131	CDNA #71 encoding
12	33.2	31.9	659	24	ABK44112	CDNA #52 encoding
13	33.2	31.9	780	24	ABK44065	CDNA #5 encoding h
14	33.2	31.9	823	24	ABK44088	CDNA #28 encoding
15	33.2	31.9	994	22	AAF44902	Human breast cance
16	33.2	31.9	1881	18	AAT59892	Coding sequence fo
17	33.2	31.9	4428	22	AAD06574	Bovine alpha(III)
18	33.2	31.9	4428	22	AAD06574	Bovine alpha(III)
19	33.2	31.9	4428	22	AAD06578	Porcine alpha(III)
20	33.2	31.9	5460	17	AAT16508	Vector pAC3A1 cont
21	33.2	31.9	5460	22	ABA83117	Collagen type III
22	33.2	31.9	5460	24	ABL92101	Human Tumour Endot
23	33.2	31.9	5466	23	AAS79378	DNA encoding novel
24	32.8	31.5	501	13	AAQ26041	Human collagen XI
25	32.8	31.5	549	24	ABK44109	CDNA #49 encoding
26	32.8	31.5	1121	23	AAS89670	DNA encoding novel
27	32.8	31.5	5010	24	ABL92114	Human Tumour Endot
28	32.8	31.5	6158	24	ABL62095	Colon adenocarcino
29	32.8	31.5	6158	24	ABL62095	Lung cancer relate
30	32.8	31.5	6158	24	ABK35486	Human endometrial
31	32.8	31.5	6436	22	AAI60019	Human polynucleoti
32	32.8	31.5	51935	22	AAK75883	Human immune/haema
33	32.4	31.2	886	24	ABK44084	CDNA #4 encoding h
34	32.4	31.2	1317	22	AAH48067	Murine HSP47 inter
35	32	30.8	61710	22	AAK83782	Human immune/haema
36	31.2	30.0	1619	22	AAS27365	CDNA encoding nove
37	31.2	30.0	1619	22	AAS34830	CDNA encoding nove
38	31.2	30.0	6674	22	AAI58233	Human polynucleoti
39	31	29.8	3226	24	ABR84606	Human CDNA differe
40	31	29.8	3226	24	ABR84606	Lung cancer relate
41	30.4	29.2	707	22	AAS44890	Human contig polyn
42	30.4	29.2	821	24	ABL51013	Human EDAL-II exon
43	30.4	29.2	1176	24	ABL51020	Human EDAL-II open
44	30.4	29.2	1574	24	ABL51009	Human EDAL-II enco
45	30.4	29.2	2686	22	AAS27013	CDNA encoding nove

## ALIGNMENTS

RESULT 1  
AAL44066  
ID AAL44066 standard; CDNA; 1338 BP.

XX AC AAL44066;

XX DT 27-SEP-2002 (first entry)

XX DE Human genset metabolic gene (GMG-9) CDNA sequence.

XX KW Human; gene; ss; gene therapy; genset metabolic gene; GMG-7A; GMG-7B;  
GMG-8; GMG-9; GMG-10; GMG-11; metabolic-related disorder; obesity;  
XX KW impaired glucose tolerance; insulin resistance; Syndrome X;  
XX KW Type II diabetes; hyperlipidaemia; atherosclerosis; hypertension;  
XX KW heart disease; cardiac insufficiency; coronary insufficiency;  
XX KW high blood pressure; insulin sensitiser;  
XX KW non-insulin dependent diabetes mellitus.

XX OS Homo sapiens.

XX Key Location/Qualifiers  
XX CDS 1..1338  
XX FT /\*tag= a  
XX FT /partial  
XX FT /product= "Human GMG-9 protein"  
XX FT /note= "No stop codon is given"

PN WO200255694-A2.

PD 18-JUL-2002.

XX 15-JAN-2002; 2002WO-IB01215.

```
XX 16-JAN-2001; 2001US-262235P.
XX (GEST ) GENSET.
XX Erickson MR, Bour BA, Bihain B, Tanaka H;
XX WPI; 2002-557821/59.
XX P-PSDB; AAO15423.
XX Treating or preventing a metabolic-related disease or disorder, e.g.
XX obesity, impaired glucose tolerance, insulin resistance, Syndrome X, or
XX Type II diabetes, comprises administering Genset Metabolic Genes -
XX Disclosure: Page 122-124; 128pp; English.
XX The invention comprises the amino acid and coding sequences of six human
XX genset metabolic genes (GMG-7A, GMG-7B, GMG-8, GMG-9, GMG-10 and GMG-11).
XX The GMG DNA and protein sequences of the invention are useful for
XX treating or preventing metabolic-related disorders, such as: obesity;
XX impaired glucose tolerance; insulin resistance; Syndrome X; Type II
XX diabetes; hyperlipidaemia; atherosclerosis; hypertension; and heart
XX diseases (e.g. cardiac insufficiency, coronary insufficiency or high
XX blood pressure). The GMG DNA and protein sequences of the invention may
XX also be used as insulin sensitizers - for improving insulin sensitivity
XX in persons with non-insulin dependent diabetes mellitus. The present cDNA
XX sequence encodes the human GMG-9 protein.
XX Sequence 1338 BP; 359 A; 288 C; 350 G; 341 T; 0 other;
Query Match 56.7%; Score 59; DB 24; Length 1338;
Best Local Similarity 92.5%; Pred. NO. 2.2e-09;
Matches 62; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 38 TGGGGCCCTCGTGCCACCAGGTTTACCACAATATACAGGAGAAATAAGTGAATGACAA 97
Db 50 TGGATGTCCTGGTGGCCCGAGTTTACCACAATATACAGGAGAAATAAGTGAATGACAA 109
QY 98 AATGCC 104
Db 110 AATGCC 116
RESULT 2
ABI99819 standard; cDNA; 6512 BP.
AC ABI99819;
XX Mouse ischaemic condition related cDNA sequence SEQ ID NO:932.
DE 07-MAR-2002 (first entry)
XX Mouse ischaemic condition related cDNA sequence SEQ ID NO:932.
KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
KW vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
XX Mus musculus.
XX WO200188188-A2.
XX 22-NOV-2001.
XX 18-MAY-2001; 2001WO-JP04192.
XX 18-MAY-2000; 2000JP-0145977.
XX (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
XX WPI; 2002-034733/04.
XX P-PSDB; ABB57334.
XX
```

```
PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT expression levels of particular genes defined in the specification or
PT by determining the expression profile of a gene group comprising these
PT genes -
XX Claim 2; Page 2340-2352; 2690pp; English.
XX The present invention describes a method for examining ischaemic
XX conditions, comprising measuring the expression levels of particular
XX genes (I) in a test sample or determining the expression profile of a
XX gene group in the sample comprising genes selected from (I). The method
XX is useful for examining the ischaemic condition (e.g. compressive
XX ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
XX expression levels of particular genes (ABI99202 to ABI99912, encoding
XX the protein sequences in ABB57020 to ABB57374) or by determining the
XX expression profile of a gene group comprising these genes. The
XX expression levels or expression profiles produced by these genes are
XX used as an indicator when screening for ischaemic condition-improving
XX drugs or therapeutics for ischaemic diseases. ABI99913 and ABI99914
XX represent PCR primers for a mouse ischaemic condition related sequence,
XX which are used in the exemplification of the present invention.
XX Sequence 6512 BP; 1553 A; 1798 C; 1950 G; 1211 T; 0 other;
Query Match 35.4%; Score 36.8; DB 24; Length 6512;
Best Local Similarity 61.5%; Pred. NO. 0.065;
Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
QY 2 CTGGACCTCCAGCACACACCCAGCCGCCACAGAGAGTGGGGCCTCTGTGCACCAGGTT 61
Db 2692 CAGGACAGTCAGGCCCTCCCTGGCCCTCTGGACAGCAGGGGACACCTGGAGTTCAGG 2751
QY 62 TACCACAATATACAGGAGAAATAAGTGAATGACAA 97
Db 2752 TCCACAGTCTAAAGTGAAATGGGTGTCATGGGAA 2787
RESULT 3
AAH98343 standard; cDNA; 5467 BP.
XX AAH98343;
XX AC AAH98343;
XX 12-OCT-2001 (first entry)
XX Human EST-derived coding sequence SEQ ID NO: 200.
XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; ss.
XX Homo sapiens.
XX WO200154477-A2.
XX 02-AUG-2001.
XX 25-JAN-2001; 2001WO-US02687.
XX 25-JAN-2000; 2000US-0491404.
XX 17-JUL-2000; 2000US-0617746.
XX 03-AUG-2000; 2000US-0631451.
XX 15-SEP-2000; 2000US-0663870.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
XX Cao Y, Drmanac RA, Zhang J, Werhman T;
XX WPI; 2001-476164/51.
XX P-PSDB; AAM23684.
XX
```

PT Isolated polypeptide for treatment of diseases, diagnostics, raising  
 PT antibodies and research use -  
 XX Claim 1; Page 314-315; 1275pp; English.  
 PS  
 CC The present invention provides the protein and coding sequences of novel  
 CC proteins from a variety of organisms, including human, dog, cat, horse,  
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
 CC from the organism of interest. They can be used in diagnostics,  
 CC forensics, gene mapping, identification of mutations, to assess  
 CC biodiversity and for nutritional purposes. The present sequence is a cDNA  
 CC of the invention.  
 XX  
 SQ Sequence 5467 BP; 1330 A; 1323 C; 1550 G; 1264 T; 0 other;  
 Query Match 32.7%; Score 34; DB 22; Length 5467;  
 Best Local Similarity 61.1%; Pred. No. 0.5;  
 Matches 55; Conservative 0; Mismatches 35; Indels 0; Gaps 0;  
 QY 1 GCTGGACCTCCAGCACACCCAGGCCGCCAGGAAGTGGGCCCTCTGGTGCACACGAGT 60  
 Db 631 GCTGCCCCGCCAGGCCGCCGCCGCCCTGTACATCTGGTTCATCTGGTTCCTCCCTGGA 690  
 QY 61 TTACCACAAATATACAGGAGAAATAAGTGAA 90  
 Db 691 TCTCAGGATACCAAGAGGCCCTGGTGAA 720  
 RESULT 4  
 AAH98411  
 ID AAH98411 standard; cDNA; 5468 BP.  
 XX  
 AC AAH98411;  
 DT 12-OCT-2001 (first entry)  
 XX  
 DE Human EST-derived coding sequence SEQ ID NO: 268.  
 XX  
 KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
 KW diagnostics; forensic test; gene mapping; genetic disorder;  
 KW biodiversity; gene therapy; nutrition; ss.  
 XX Homo sapiens.  
 OS  
 PN WO200154477-A2.  
 PD 02-AUG-2001.  
 XX  
 PF 25-JAN-2001; 2001WO-US02687.  
 XX  
 PR 25-JAN-2000; 2000US-0491404.  
 PR 17-JUL-2000; 2000US-0617746.  
 PR 03-AUG-2000; 2000US-0631451.  
 PR 15-SEP-2000; 2000US-0663870.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
 PI Cao Y, Drmanac RA, Zhang J, Werhman T;  
 XX  
 DR WPI; 2001-476164/51.  
 DR P-PSDB; AAM23752.  
 XX  
 XX Isolated polypeptide for treatment of diseases, diagnostics, raising  
 PT antibodies and research use -  
 XX  
 PS Claim 1; Page 381-382; 1275pp; English.  
 XX  
 CC The present invention provides the protein and coding sequences of novel  
 CC proteins from a variety of organisms, including human, dog, cat, horse,  
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea

CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
 CC from the organism of interest. They can be used in diagnostics,  
 CC forensics, gene mapping, identification of mutations, to assess  
 CC biodiversity and for nutritional purposes. The present sequence is a cDNA  
 CC of the invention.  
 XX  
 SQ Sequence 5468 BP; 1330 A; 1323 C; 1551 G; 1264 T; 0 other;  
 Query Match 32.7%; Score 34; DB 22; Length 5468;  
 Best Local Similarity 61.1%; Pred. No. 0.5;  
 Matches 55; Conservative 0; Mismatches 35; Indels 0; Gaps 0;  
 QY 1 GCTGGACCTCCAGCACACCCAGGCCGCCAGGAAGTGGGCCCTCTGGTGCACACGAGT 60  
 Db 631 GCTGCCCCGCCAGGCCGCCGCCGCCCTGTACATCTGGTTCATCTGGTTCCTCCCTGGA 690  
 QY 61 TTACCACAAATATACAGGAGAAATAAGTGAA 90  
 Db 691 TCTCAGGATACCAAGAGGCCCTGGTGAA 720  
 RESULT 5  
 ABK44107  
 ID ABK44107 standard; cDNA; 561 BP.  
 XX  
 AC ABK44107;  
 XX  
 DT 21-MAY-2002 (first entry)  
 XX  
 DE cDNA #47 encoding human pancreatic tumour protein.  
 XX  
 KW Human; pancreatic tumour protein; immune response; pancreatic cancer;  
 KW development of cancer; cancer progression; cytostatic; gene; ss.  
 XX Homo sapiens.  
 OS  
 PN WO200212331-A2.  
 PD 14-FEB-2002.  
 XX  
 PF 06-AUG-2001; 2001WO-US24619.  
 XX  
 PR 07-AUG-2000; 2000US-223130P.  
 PR 30-JAN-2001; 2001US-265447P.  
 PR 15-MAY-2001; 2001US-291201P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Pyle RA, Xu J, Kalos MD;  
 XX  
 DR WPI; 2002-241741/29.  
 XX  
 PT Novel polynucleotide encoding pancreatic tumour polypeptides, useful in  
 PT pharmaceutical compositions, e.g. vaccines, for treating pancreatic  
 PT cancers -  
 XX  
 PS Claim 1; Page 126; 167pp; English.  
 XX  
 CC The present invention relates to the isolation of cDNA sequences  
 CC encoding human pancreatic tumour proteins. The polynucleotide  
 CC sequences encoding human pancreatic tumour proteins are useful for  
 CC stimulating an immune response in a patient and treating pancreatic  
 CC cancer in a patient. A host cell that expresses these polynucleotides  
 CC is useful for determining the presence of cancer in a patient. A  
 CC composition comprising the polynucleotide, its encoded protein, or an  
 CC antibody that binds to the protein may be used in the diagnosis,  
 CC prevention and/or treatment of diseases, particularly pancreatic  
 CC cancer. The sequences of the invention are also useful in pharmaceutical  
 CC compositions, e.g. vaccines, for the diagnosis and treatment of  
 CC pancreatic cancer. Such compositions may be useful for inhibiting the  
 CC development of cancer in a patient, or as markers for the progression  
 CC of cancer. The polynucleotide sequences may also be used as probes  
 CC or primers for nucleic acid hybridisation assays. ABK44061-ABK44209



QY 62 TACCACAATATACAGGAGAAATAAGT 87  
 |||| | | | | | | | |  
 Db 464 CTCGAGGACCTCCTGGTGCTATAGGT 489

RESULT 9	
ABK44128	
ID	ABK44128 standard; cDNA; 655 BP.
AC	
XX	ABK44128;
XX	
DT	21-MAY-2002 (first entry)
XX	
XX	
DE	cDNA #68 encoding human pancreatic tumour protein.

XX Homo sapiens.  
XX  
XX WO200212331-A2.  
XX PN  
XX PD  
PD 14-FEB-2002.

XX	06-AUG-2001; 2001WO-US24619.
PF	
XX	
PR	07-AUG-2000; 2000US-223130P.
PR	30-JAN-2001; 2001US-265447P.
PR	15-MAY-2001; 2001US-291201P.
XX	
PA	(CORI-) CORIYA CORP.
XX	
PI	Pyle RA, Xu J, Kalos MD;
XX	
XX	WPI: 2002-241741/29.

Novel polynucleotide encoding pancreatic tumour polypeptides, useful in pharmaceutical compositions, e.g. vaccines, for treating pancreatic cancers -

Claim 1; Page 132; 167pp; English.

The present invention relates to the isolation of cDNA sequences encoding human pancreatic tumour proteins. The polynucleotide sequences encoding human pancreatic tumour proteins are useful for stimulating an immune response in a patient and treating pancreatic cancer in a patient. A host cell that expresses these polynucleotides is useful for determining the presence of cancer in a patient. A composition comprising the polynucleotide, its encoded protein, or an antibody that binds to the protein may be used in the diagnosis, prevention and/or treatment of diseases, particularly pancreatic cancer. The sequences of the invention are also useful in pharmaceutical compositions, e.g. vaccines, for the diagnosis and treatment of pancreatic cancer. Such compositions may be useful for inhibiting the development of cancer in a patient, or as markers for the progression of cancer. The polynucleotide sequences may also be used as probes or primers for nucleic acid hybridisation assays. ABK44061-ABK44209 represent cDNA sequences encoding for human pancreatic tumour proteins.

Sequence 655 BP: 150 A: 189 C: 171 G: 144 T: 1 other:

Query Match	31.9%	Score 33.2;	DB 24;	Length 655;
Best Local Similarity	61.6%	Pred. No. 0.44;		
Matches 53;	Conservative	0;	Mismatches 33;	Indels 0;
Matches 53;	Conservative	0;	Mismatches 33;	Indels 0;

Qy	2	CTGACCTTCAGACACACCCAGCGCCCTCCCTGGTGCACCAAGTTT	61
Db	404	CTGGATCTCCAGGATACCAAGGACCCCTGTTGAACCTGGCAAGCTGGTCTTCAGGCC	463
Qy	62	TACCACAATATACAGGAGAATAAGT	87
Db	464	CTCCAGGACCTCCTGGTGCCTATAGGT	489

```
RESULT 10
ABK44129
ID ABK44129 standard; cDNA; 656 BP.
XX
AC ABK44129;
XX
XX
DT 21-MAY-2002 (first entry)
DE
DE
KW Human; pancreatic tumour protein; immune response; pancreatic cancer;
KW development of cancer; cancer progression; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200212331-A2.
XX
PD 14-FEB-2002.
XX
PF 06-AUG-2001; 2001WO-US24619.
XX
PR 07-AUG-2000; 2000US-223130P.
PR 30-JAN-2001; 2001US-265447P.
PR 15-MAY-2001; 2001US-291201P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Pyle RA, Xu J, Kalos MD;
XX
PI WO2002-241741/29.
XX
DR
PT Novel polynucleotide encoding pancreatic tumour polypeptides, useful in
PT pharmaceutical compositions, e.g. vaccines, for treating pancreatic
PT cancers -
XX
XX
PS Claim 1; Page 133; 167pp; English.
XX
CC The present invention relates to the isolation of cDNA sequences
CC encoding human pancreatic tumour proteins. The polynucleotide
CC sequences encoding human pancreatic tumour proteins are useful for
CC stimulating an immune response in a patient and treating pancreatic
CC cancer in a patient. A host cell that expresses these polynucleotides
CC is useful for determining the presence of cancer in a patient. A
CC composition comprising the polynucleotide, its encoded protein, or an
CC antibody that binds to the protein may be used in the diagnosis,
CC prevention and/or treatment of diseases, particularly pancreatic
CC cancer. The sequences of the invention are also useful for inhibiting the
CC development of cancer in a patient, or as markers for the progression
CC of cancer. The polynucleotide sequences may also be used as probes
CC or primers for nucleic acid hybridisation assays. ABK44061-ABK44209
CC represent cDNA sequences encoding for human pancreatic tumour proteins.
XX
SQ Sequence 656 BP; 152 A; 188 C; 168 G; 145 T; 3 other;

Query Match 31.9%; Score 33.2; DB 24; Length 656;
Best Local Similarity 61.6%; Pred. No. 0.44;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 2 CTGGACCTCCAGCACACCCAGGCCCGCCAGAGAGTGGGGCTCCTGGTGACCAAGGTT 61
DB 404 CTGGATCTCCAGGATACCAAGACCCCTGGTGAACCTGGGCAAGCTGGTCTTCAGGCC 463

QY 62 TACCACATATACAGGAGAAATAAGT 87
DB 464 CTCGAGGACCTCCTGGTGCTATAGT 489

RESULT 11
ABK44131
ID ABK44131 standard; cDNA; 658 BP.
XX
AC ABK44131;
XX
XX
DT 21-MAY-2002 (first entry)
DE
DE
KW Human; pancreatic tumour protein; immune response; pancreatic cancer;
KW development of cancer; cancer progression; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200212331-A2.
XX
PD 14-FEB-2002.
XX
PF 06-AUG-2001; 2001WO-US24619.
XX
PR 07-AUG-2000; 2000US-223130P.
PR 30-JAN-2001; 2001US-265447P.
PR 15-MAY-2001; 2001US-291201P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Pyle RA, Xu J, Kalos MD;
XX
PI WO2002-241741/29.
XX
DR
PT Novel polynucleotide encoding pancreatic tumour polypeptides, useful in
PT pharmaceutical compositions, e.g. vaccines, for treating pancreatic
PT cancers -
XX
XX
PS Claim 1; Page 133; 167pp; English.
XX
CC The present invention relates to the isolation of cDNA sequences
CC encoding human pancreatic tumour proteins. The polynucleotide
CC sequences encoding human pancreatic tumour proteins are useful for
CC stimulating an immune response in a patient and treating pancreatic
CC cancer in a patient. A host cell that expresses these polynucleotides
CC is useful for determining the presence of cancer in a patient. A
CC composition comprising the polynucleotide, its encoded protein, or an
CC antibody that binds to the protein may be used in the diagnosis,
CC prevention and/or treatment of diseases, particularly pancreatic
CC cancer. The sequences of the invention are also useful in pharmaceutical
CC compositions, e.g. vaccines, for the diagnosis and treatment of
CC pancreatic cancer. Such compositions may be useful for inhibiting the
CC development of cancer in a patient, or as markers for the progression
CC of cancer. The polynucleotide sequences may also be used as probes
CC or primers for nucleic acid hybridisation assays. ABK44061-ABK44209
CC represent cDNA sequences encoding for human pancreatic tumour proteins.
XX
SQ Sequence 658 BP; 151 A; 190 C; 172 G; 144 T; 1 other;
```

```
XX
AC ABK44131;
XX
XX
DT 21-MAY-2002 (first entry)
DE
DE
KW Human; pancreatic tumour protein; immune response; pancreatic cancer;
KW development of cancer; cancer progression; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200212331-A2.
XX
PD 14-FEB-2002.
XX
PF 06-AUG-2001; 2001WO-US24619.
XX
PR 07-AUG-2000; 2000US-223130P.
PR 30-JAN-2001; 2001US-265447P.
PR 15-MAY-2001; 2001US-291201P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Pyle RA, Xu J, Kalos MD;
XX
PI WO2002-241741/29.
XX
DR
PT Novel polynucleotide encoding pancreatic tumour polypeptides, useful in
PT pharmaceutical compositions, e.g. vaccines, for treating pancreatic
PT cancers -
XX
XX
PS Claim 1; Page 133; 167pp; English.
XX
CC The present invention relates to the isolation of cDNA sequences
CC encoding human pancreatic tumour proteins. The polynucleotide
CC sequences encoding human pancreatic tumour proteins are useful for
CC stimulating an immune response in a patient and treating pancreatic
CC cancer in a patient. A host cell that expresses these polynucleotides
CC is useful for determining the presence of cancer in a patient. A
CC composition comprising the polynucleotide, its encoded protein, or an
CC antibody that binds to the protein may be used in the diagnosis,
CC prevention and/or treatment of diseases, particularly pancreatic
CC cancer. The sequences of the invention are also useful in pharmaceutical
CC compositions, e.g. vaccines, for the diagnosis and treatment of
CC pancreatic cancer. Such compositions may be useful for inhibiting the
CC development of cancer in a patient, or as markers for the progression
CC of cancer. The polynucleotide sequences may also be used as probes
CC or primers for nucleic acid hybridisation assays. ABK44061-ABK44209
CC represent cDNA sequences encoding for human pancreatic tumour proteins.
XX
SQ Sequence 658 BP; 151 A; 190 C; 172 G; 144 T; 1 other;

Query Match 31.9%; Score 33.2; DB 24; Length 658;
Best Local Similarity 61.6%; Pred. No. 0.44;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 2 CTGGACCTCCAGCACACCCAGGCCCGCCAGAGAGTGGGGCTCCTGGTGACCAAGGTT 61
DB 404 CTGGATCTCCAGGATACCAAGACCCCTGGTGAACCTGGGCAAGCTGGTCTTCAGGCC 463

QY 62 TACCACATATACAGGAGAAATAAGT 87
DB 464 CTCGAGGACCTCCTGGTGCTATAGT 489

RESULT 12
ABK44112
ID ABK44112 standard; cDNA; 659 BP.
XX
AC ABK44112;
XX
XX
DT 21-MAY-2002 (first entry)
```

XX cDNA #52 encoding human pancreatic tumour protein.  
DE Human; pancreatic tumour protein; immune response; pancreatic cancer;  
KW development of cancer; cancer progression; cytostatic; gene; ss.  
KW Homo sapiens.  
XX WO200212331-A2.  
XX 14-FEB-2002.  
XX 06-AUG-2001; 2001WO-US24619.  
XX 07-AUG-2000; 2000US-223130P.  
XX 30-JAN-2001; 2001US-265447P.  
XX 15-MAY-2001; 2001US-291201P.  
XX (CORI-) CORIXA CORP.  
XX Pyle RA, Xu J, Kalos MD;  
XX WPI; 2002-241741/29.  
XX Novel polynucleotide encoding pancreatic tumour polypeptides, useful in  
PT pharmaceutical compositions, e.g. vaccines, for treating pancreatic  
PT cancers -  
XX Claim 1; Page 127-128; 167pp; English.  
XX The present invention relates to the isolation of cDNA sequences  
CC encoding human pancreatic tumour proteins. The polynucleotide  
CC sequences encoding human pancreatic tumour proteins are useful for  
CC stimulating an immune response in a patient and treating pancreatic  
CC cancer in a patient. A host cell that expresses these polynucleotides  
CC is useful for determining the presence of cancer in a patient. A  
CC antibody that binds to the protein may be used in the diagnosis,  
CC prevention and/or treatment of diseases, particularly pancreatic  
CC cancer. The sequences of the invention are also useful in pharmaceutical  
CC compositions, e.g. vaccines, for the diagnosis and treatment of  
CC pancreatic cancer. Such compositions may be useful for inhibiting the  
CC development of cancer in a patient, or as markers for the progression  
CC of cancer. The polynucleotide sequences may also be used as probes  
CC or primers for nucleic acid hybridisation assays. ABK44061-ABK44209  
CC represent cDNA sequences encoding for human pancreatic tumour proteins.  
XX Sequence 659 BP; 152 A; 189 C; 174 G; 143 T; 1 other;  
SQ  
Query Match 31.9%; Score 33.2; DB 24; Length 659;  
Best Local Similarity 61.6%; Pred. No. 0.44;  
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
QY 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAAGTGGGCCCTCTGGTGCCACCAAGTT 61  
DB 404 CTGGATCTCCAGGATACCAAGGACCCCTGGTGAACCTGGGCAAGCTGCTTCAGGCC 463  
QY 62 TACCACAAATATACAGGAGAAATAAGT 87  
DB 464 CTCAGGACCTCTGGTGCTATAGTT 489  
RESULT 13  
ID ABK44065/C  
XX ABK44065 standard; cDNA; 780 BP.  
XX AC ABK44065;  
XX 21-MAY-2002 (first entry)  
DE cDNA #5 encoding human pancreatic tumour protein.  
XX cDNA #5 encoding human pancreatic tumour protein.  
KW Human; pancreatic tumour protein; immune response; pancreatic cancer;

KW development of cancer; cancer progression; cytostatic; gene; ss.  
XX Homo sapiens.  
XX WO200212331-A2.  
XX 14-FEB-2002.  
XX 06-AUG-2001; 2001WO-US24619.  
XX 07-AUG-2000; 2000US-223130P.  
XX 30-JAN-2001; 2001US-265447P.  
XX 15-MAY-2001; 2001US-291201P.  
XX (CORI-) CORIXA CORP.  
XX Pyle RA, Xu J, Kalos MD;  
XX WPI; 2002-241741/29.  
XX Novel polynucleotide encoding pancreatic tumour polypeptides, useful in  
PT pharmaceutical compositions, e.g. vaccines, for treating pancreatic  
PT cancers -  
XX Claim 1; Page 111; 167pp; English.  
XX The present invention relates to the isolation of cDNA sequences  
CC encoding human pancreatic tumour proteins. The polynucleotide  
CC sequences encoding human pancreatic tumour proteins are useful for  
CC stimulating an immune response in a patient and treating pancreatic  
CC cancer in a patient. A host cell that expresses these polynucleotides  
CC is useful for determining the presence of cancer in a patient. A  
CC antibody that binds to the protein may be used in the diagnosis,  
CC prevention and/or treatment of diseases, particularly pancreatic  
CC cancer. The sequences of the invention are also useful in pharmaceutical  
CC compositions, e.g. vaccines, for the diagnosis and treatment of  
CC pancreatic cancer. Such compositions may be useful for inhibiting the  
CC development of cancer in a patient, or as markers for the progression  
CC of cancer. The polynucleotide sequences may also be used as probes  
CC or primers for nucleic acid hybridisation assays. ABK44061-ABK44209  
CC represent cDNA sequences encoding for human pancreatic tumour proteins.  
XX Sequence 780 BP; 163 A; 211 C; 218 G; 171 T; 17 other;  
SQ  
Query Match 31.9%; Score 33.2; DB 24; Length 780;  
Best Local Similarity 61.6%; Pred. No. 0.47;  
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
QY 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAAGTGGGCCCTCTGGTGCCACCAAGTT 61  
DB 392 CTGGATCTCCAGGATACCAAGGACCCCTGGTGAACCTGGGCAAGCTGCTTCAGGCC 333  
QY 62 TACCACAAATATACAGGAGAAATAAGT 87  
DB 332 CTCAGGACCTCTGGTGCTATAGTT 307  
RESULT 14  
ID ABK44088  
XX ABK44088 standard; cDNA; 823 BP.  
XX AC ABK44088;  
XX 21-MAY-2002 (first entry)  
DE cDNA #28 encoding human pancreatic tumour protein.  
XX Human; pancreatic tumour protein; immune response; pancreatic cancer;  
KW development of cancer; cancer progression; cytostatic; gene; ss.  
XX Homo sapiens.  
XX

```
PN WO200212331-A2.
XX
PD 14-FEB-2002.
XX
PF 06-AUG-2001; 2001WO-US24619.
XX
PR 07-AUG-2000; 2000US-223130P.
PR 30-JAN-2001; 2001US-265447P.
PR 15-MAY-2001; 2001US-291201P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Pyle RA, Xu J, Kalos MD;
XX
DR WPI; 2002-241741/29.
XX
PT Novel polynucleotide encoding pancreatic tumour polypeptides, useful in
PT pharmaceutical compositions, e.g. vaccines, for treating pancreatic
PT cancers.
XX
PS Claim 1; Page 119; 167pp; English.
XX
CC The present invention relates to the isolation of cDNA sequences
CC encoding human pancreatic tumour proteins. The polynucleotide
CC sequences encoding human pancreatic tumour proteins are useful for
CC stimulating an immune response in a patient and treating pancreatic
CC cancer in a patient. A host cell that expresses these polynucleotides
CC is useful for determining the presence of cancer in a patient. A
CC composition comprising the polynucleotide, its encoded protein, or an
CC antibody that binds to the protein may be used in the diagnosis,
CC prevention and/or treatment of diseases, particularly pancreatic
CC cancer. The sequences of the invention are also useful in pharmaceutical
CC compositions, e.g. vaccines, for the diagnosis and treatment of
CC pancreatic cancer. Such compositions may be useful for inhibiting the
CC development of cancer in a patient, or as markers for the progression
CC of cancer. The polynucleotide sequences may also be used as probes
CC or primers for nucleic acid hybridisation assays. ABK44061-ABK44209
XX represent cDNA sequences encoding for human pancreatic tumour proteins.
XX
SQ Sequence 823 BP; 182 A; 228 C; 237 G; 173 T; 3 other;

Query Match 31.9%; Score 33.2; DB 24; Length 823;
Best Local Similarity 61.6%; Pred. No. 0.48;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCTGTGCACCAGGTT 61
DB 413 CTGGATCTCCAGGATACCAAGGACCCCTCGTGAACCTGGGCAAGCTGCTCTCAGGCC 472

QY 62 TACCACAATATACAGGAGAAATAAGT 87
DB 473 CTCAGGACCTCTCTGTGCTATAGGT 498

RESULT 15
AAF44902
ID AAF44902 standard; cDNA; 994 BP.
XX
AC AAF44902;
XX
DT 28-MAR-2001 (first entry)
XX
DE Human breast cancer related protein coding sequence SEQ ID NO: 58.
XX
KW Human; breast cancer; diagnosis; therapy; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200078960-A2.
XX
PD 28-DEC-2000.
XX
PF 23-JUN-2000; 2000WO-US17536.
XX

XX
PR 23-JUN-1999; 99US-0140903.
PR 12-OCT-1999; 99US-0158980.
XX
PA (CORI-) CORIXA CORP.
XX
PI Yuqiu J, Mitcham JL;
XX
DR WPI; 2001-041426/05.
XX
PT New polynucleotides encoding breast tumor specific proteins, useful for
PT prevention, treatment and diagnosis of breast cancer -
XX
PS Claim 25; Page 138; 165pp; English.
XX
CC The present invention provides the coding sequences for a number of
CC breast cancer related proteins. These can be used in vaccinations
CC against, diagnosis of and treatment of cancer, particularly breast
CC cancer.
XX
SQ Sequence 994 BP; 234 A; 260 C; 262 G; 237 T; 1 other;

Query Match 31.9%; Score 33.2; DB 22; Length 994;
Best Local Similarity 61.6%; Pred. No. 0.51;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCTGTGCACCAGGTT 61
DB 685 CTGGATCTCCAGGATACCAAGGACCCCTCGTGAACCTGGGCAAGCTGCTCTCAGGCC 744

QY 62 TACCACAATATACAGGAGAAATAAGT 87
DB 745 CTCAGGACCTCTCTGTGCTATAGGT 770

Search completed: February 19, 2003, 22:56:18
Job time : 40.9937 secs
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:16:26 ; Search time 331.581 Seconds  
(without alignments)  
9128.055 Million cell updates/sec

Title: US-09-997-610-1\_COPY\_53\_156

Perfect score: 104

Sequence: 1 gctgacacctccagcacacc.....agtgaatgacaaaatgcc 104

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vi:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_mus:\*
- 34: em\_htg\_pln:\*
- 35: em\_htg\_rod:\*
- 36: em\_htg\_man:\*
- 37: em\_htg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_htgo\_hum:\*
- 40: em\_htgo\_mus:\*
- 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60	57.7	145880	9	HS302D9
2	40.6	39.0	5851	3	AF282902
3	37.4	36.0	287	3	AF282902 Hydra vul
4	36.8	35.4	3612	10	OOS420010
5	36.8	35.4	6512	6	MMCOLA4
6	36.8	35.4	6512	6	AX306181
7	36.2	34.8	694	3	HSNCOL1
8	36.2	34.8	36532	3	CEF57B1
9	36.2	34.8	289619	3	CEY51H4A
10	35.6	34.2	11942	1	AE010554
11	35.2	33.8	1993	3	BMCOIGMR
12	34.8	33.5	3734	14	HVSX99519
13	34.8	33.5	6522	3	U58736
14	34.2	32.9	31110	3	CBRG14K24
15	34	32.7	3234	9	HSC3A1R
16	34	32.7	87849	3	AC084453
17	33.8	32.5	1776	10	RNU57362
18	33.4	32.1	97083	9	AC010289
19	33.4	32.1	107655	2	AC084840
20	33.4	32.1	147416	9	AC005549
21	33.4	32.1	156000	2	AC060815
22	33.4	32.1	159694	2	AC098586
23	33.4	32.1	162214	9	AC093809
24	33.4	32.1	177941	9	AL445423
25	33.4	32.1	188460	9	AC010542
26	33.4	32.1	189876	2	AC084810
27	33.4	32.1	316776	2	AL603862
28	33.2	31.9	994	6	AX067354
29	33.2	31.9	1881	6	A60690
30	33.2	31.9	1881	6	AR123971
31	33.2	31.9	3902	9	BC028178
32	33.2	31.9	4428	6	AX146422
33	33.2	31.9	4428	6	AX146424
34	33.2	31.9	4428	6	AX146430
35	33.2	31.9	5460	6	AX302553
36	33.2	31.9	5460	6	AX393295
37	33.2	31.9	5460	6	E10600
38	33.2	31.9	5460	9	HSCOL3A1
39	33.2	31.9	68808	2	AC022708
40	33.2	31.9	96799	9	AC104647
41	33.2	31.9	105936	3	AC084440
42	33.2	31.9	198491	2	AC022249
43	33	31.7	2247	3	CBU84501
44	33	31.7	4235	10	BC016479
45	33	31.7	9448	3	CEU22327

ALIGNMENTS

RESULT 1

HS302D9

LOCUS

DEFINITION

Human DNA sequence from clone Rpl-302D9 on chromosome 22 Contains

GSSS, complete sequence.

ACCESSION 282198

VERSION 282198.2

KEYWORDS HTG.

SOURCE HTG.

ORGANISM Homo sapiens.

REFERENCE 1 (bases 1 to 145880)

AUTHORS Bridgeman,A.

TITLE Direct Submission

HS302D9 145880 bp DNA linear PRI 12-DEC-1999  
Human DNA sequence from clone Rpl-302D9 on chromosome 22 Contains  
GSSS, complete sequence.

282198

282198.2

GI:6572207

HTG.

HTG.

Homo sapiens.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 145880)

Bridgeman,A.

Direct Submission

## JOURNAL

Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

## COMMENT

On Dec 13, 1999 this sequence version replaced gi:3164067. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr22>  
 RPI-302D9 is from the library RPCI-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/> VECTOR: pCYPAC2

This sequence is the entire insert of clone RPI-302D9 The true left end of clone CTA-282F2 is at 69682 in this sequence. The true right end of clone CTA-415G2 is at 55167 in this sequence.

## FEATURES

Location/Qualifiers

source

1..145880  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="22"  
 /clone="RPI-302D9"  
 /clone\_lib="RPCI-1"

repeat\_region

188..245  
 /note="MER3 repeat: matches 144..209 of consensus"

repeat\_region

246..571  
 /note="AluX repeat: matches 1..312 of consensus"

repeat\_region

572..759  
 /note="MER3 repeat: matches 1..144 of consensus"

repeat\_region

783..933  
 /note="MERSA repeat: matches 26..187 of consensus"

repeat\_region

1033..1336  
 /note="AluSp repeat: matches 1..299 of consensus"

repeat\_region

1450..1583  
 /note="MIR repeat: matches 24..160 of consensus"

repeat\_region

1687..1752  
 /note="L2 repeat: matches 2593..2661 of consensus"

repeat\_region

2350..2660  
 /note="AluSc repeat: matches 3..309 of consensus"

repeat\_region

2684..2981  
 /note="AluX repeat: matches 2..300 of consensus"

repeat\_region

3323..3343  
 /note="MLTIE repeat: matches 116..136 of consensus"

repeat\_region

3344..3652  
 /note="AluY repeat: matches 1..309 of consensus"

repeat\_region

3653..3928  
 /note="MLTIE repeat: matches 136..359 of consensus"

repeat\_region

3929..4278  
 /note="THELB repeat: matches 3..364 of consensus"

repeat\_region

4279..4485  
 /note="MLTIE repeat: matches 359..568 of consensus"

repeat\_region

5073..5176  
 /note="52 copies 2 mer ct 78 conserved"

repeat\_region

5181..5491  
 /note="AluX repeat: matches 1..311 of consensus"

repeat\_region

6369..6485  
 /note="L2 repeat: matches 2579..2705 of consensus"

repeat\_region

6647..6685  
 /note="MADE1 repeat: matches 1..23 of consensus"

repeat\_region

6886..6987  
 /note="AluX repeat: matches 1..302 of consensus"

repeat\_region

6988..7036  
 /note="MADE1 repeat: matches 23..77 of consensus"

repeat\_region

7482..7754  
 /note="AluX repeat: matches 9..290 of consensus"

repeat\_region

7775..8060  
 /note="AluX repeat: matches 1..295 of consensus"

repeat\_region

8414..8551  
 /note="L2 repeat: matches 2553..2706 of consensus"

repeat\_region

8914..9030  
 /note="MIR repeat: matches 147..262 of consensus"

repeat\_region

9110..9280  
 /note="MIR repeat: matches 91..262 of consensus"

repeat\_region

9283..9412  
 /note="MIR repeat: matches 15..144 of consensus"

repeat\_region

9521..9679  
 /note="FAM repeat: matches 3..161 of consensus"

repeat\_region

9820..10225  
 /note="MSTB repeat: matches 2..425 of consensus"

misc\_feature

complement(10179..10678)  
 /note="match: GSS: Em:B56592"

misc\_feature

complement(10204..10728)  
 /note="match: GSS: Em:AQ701486"

misc\_feature

complement(10249..10706)  
 /note="match: GSS: Em:AQ225495"

repeat\_region

10312..10383  
 /note="MIR repeat: matches 79..150 of consensus"

misc\_feature

10718..11310  
 /note="match: GSS: Em:B14024"

misc\_feature

10784..11201  
 /note="match: GSS: Em:B43656"

repeat\_region

11838..11946  
 /note="MIR repeat: matches 20..137 of consensus"

repeat\_region

12174..12445  
 /note="L2 repeat: matches 1988..2275 of consensus"

repeat\_region

12444..12642  
 /note="MIR repeat: matches 63..241 of consensus"

misc\_feature

13017..13369  
 /note="match: STS: Em:G49301"

repeat\_region

13331..13397  
 /note="MIR repeat: matches 174..244 of consensus"

repeat\_region

13398..13698  
 /note="AluSp repeat: matches 1..302 of consensus"

repeat\_region

13699..13810  
 /note="MIR repeat: matches 76..174 of consensus"

repeat\_region

13806..13919  
 /note="MIR repeat: matches 77..189 of consensus"

repeat\_region

13945..14060  
 /note="MIR repeat: matches 24..142 of consensus"

repeat\_region

14061..14367  
 /note="AluY repeat: matches 1..301 of consensus"

repeat\_region

14368..14452  
 /note="MIR repeat: matches 141..225 of consensus"

repeat\_region

14589..14679  
 /note="MIR repeat: matches 173..262 of consensus"

misc\_feature

14597..15201  
 /note="match: GSS: Em:AQ553482"

misc\_feature

14616..15060  
 /note="match: GSS: Em:AQ370601"

repeat\_region

14868..15040  
 /note="MIR repeat: matches 49..233 of consensus"

repeat\_region

15071..15188  
 /note="L2 repeat: matches 2112..2239 of consensus"

repeat\_region

15304..15399  
 /note="MLTIB repeat: matches 1..99 of consensus"

repeat\_region

15490..15662  
 /note="AluSg1 repeat: matches 2..114 of consensus"

repeat\_region

15669..15727  
 /note="MLTIB repeat: matches 119..178 of consensus"

repeat\_region

15728..16027





```
AX306181
LOCUS AX306181 6512 bp DNA linear PAT 11-DEC-2001
DEFINITION Sequence 932 from Patent WO018188.
ACCESSION AX306181
VERSION AX306181.1 GI:17645461
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Ishikawa,K., Asai,S., Takahashi,Y., Nagata,T. and Ishii,Y.
TITLE Method for examining ischemic conditions
JOURNAL Patent: WO 018188-A 932 22-NOV-2001;
FEATURES
    Location/Qualifiers
        1..6512
            /organism="Mus musculus"
            /db_xref="taxon:10090"
BASE COUNT 1553 a 1798 c 1950 g 1211 t
ORIGIN
Query Match 35.4%; Score 36.8; DB 6; Length 6512;
Best Local Similarity 61.5%; Pred. No. 0.64;
Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
QY 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGCTCTCTGGTGCACAGGTT 61
Db 2692 CAGGACAGTCAGGCTCCCTGGCTTCTCTGGACAGGAGCACCTGGAGTTCAGGTT 2751
QY 62 TACCACAATATACAGAGAATAAGTGAATGACAA 97
Db 2752 TCCAGGTTCTAAAGGTGAATGGTGTCTATGGGAA 2787

RESULT 6
MUSCOL1A4A
LOCUS MUSCOL1A4A 6512 bp mRNA linear ROD 06-OCT-1994
DEFINITION Mus musculus alpha-1 type IV collagen (Col4a-1) mRNA, complete cds.
ACCESSION J04694
VERSION J04694.1 GI:556296
KEYWORDS alpha-1 type IV collagen.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Muthukumar,G., Blumberg,B. and Kurkinen,M.
TITLE The complete primary structure for the alpha 1-chain of mouse collagen IV. Differential evolution of collagen IV domains
JOURNAL J. Biol. Chem. 264 (11), 6310-6317 (1989)
MEDLINE 89197932
PUBMED 2703490
COMMENT On Oct 8, 1994 this sequence version replaced gi:340547.
FEATURES
    Location/Qualifiers
        1..6512
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /tissue_type="parietal endoderm"
            /dev_stage="13.5 day old embryo"
            1..6512
                /gene="Col4a-1"
                132..5141
                    /gene="Col4a-1"
                    /codon_start=1
                    /product="alpha-1 type IV collagen"
                    /protein_id="AAA50292.1"
                    /db_xref="GI:556297"
                    /translation="MGPRLSVWLLLLFALLLHEERSAAAKDCGGCGKDCDHGV
                    KQKGERGLPGQGVIGFPGMGQPGPGKQGDAGEPLGPTKGRGPCAAGYP
                    GNFLPGIPQDQGPPIPGPCNCTKGRGIPGLPGPFGFSGNPGPLPGMKMDPG
                    ETLGHVPTGLLKERGFPIGMPGSPGLPGIQGVPGFTGPGPGPPGPGKEGK
                    OMGSSFQPGKDGQGVSGVPGVPCQAQVKEKGDFAPTGEKQKGEFGVPGYGE"
gene
CDS
    132..5141
        /gene="Col4a-1"
        /codon_start=1
        /product="alpha-1 type IV collagen"
        /protein_id="AAA50292.1"
        /db_xref="GI:556297"
        /translation="MGPRLSVWLLLLFALLLHEERSAAAKDCGGCGKDCDHGV
        KQKGERGLPGQGVIGFPGMGQPGPGKQGDAGEPLGPTKGRGPCAAGYP
        GNFLPGIPQDQGPPIPGPCNCTKGRGIPGLPGPFGFSGNPGPLPGMKMDPG
        ETLGHVPTGLLKERGFPIGMPGSPGLPGIQGVPGFTGPGPGPPGPGKEGK
        OMGSSFQPGKDGQGVSGVPGVPCQAQVKEKGDFAPTGEKQKGEFGVPGYGE"

KGBPGRKQGRKPGKDEKGERGSPGIPGDSGYPLGPRQGPQGEKGEAGLPGPPGTV
IGTMPLGEKDRGYPGAPGLRGPGRKPGFPGTQCPQGPFGPTPGQAGAPGFCGERGE
KGGPFGVSLPQPSGRDAPGPPGPPGPGQPCQHTNGIIVECOPGPGDQGPCTPGQ
PGLTGEVQKQKQKESCLACTDELGRPGPGQPGGIFGPGPGAGKDRGLGCRDGL
EGLPGQGSPLIGLQPCAKGEPEIEFDMRLKGDGDPGPGPGRAGTFCGRDGH
PGLPFGPSPGISLGRGPPGVGPGSGRDIIPGPGVGPPIGVVGEKQNGAPG
GSGPLPGPKGEAGKVPLPPLPGGAAGLPGSPFGPGQDGRFPPTGPRGIPGEKG
AVQPGIGFPLGPKGVGDLGELGIPGSPGPGFNGLPNFCPQKQKGEIGLPG
LKQPGPLGIPGTPGKGSIGGPGVPEOGLTGPGLQIGRDPGPGVQGPAGPPGV
PGQPGPLGIPGTPGKGSIGGPGVPEOGLTGPGLQIGRDPGPGVQGPAGPPGV
SGPLGPGQOQTPGVPGFPGSKGEMVMTCTPGQSPGPGAGTGLPCEKGDHGLPGSS
GPRDGFGRDGDVGLPGMPGSMHVDMSKMGQKQDQGEKQIGTIDGKSGRDPG
TPGVPKDQAGHPGPGPKDPLSGTGPGLPGKPGSVGMGLPGSPGKGVPGI
PGSQGVPSPGPKGEKQKESGLGIGIPRPGDQDQGLAGFPGSPGKGEKGSAG
TPQMPSGPRGSPGNIHPSGLPGEKDKGLPGLDGVPGVKGAGLGTGPTGP
AGQKGPBGDPGIPGSAKEGQGVPRGPGFPGSKGDKGSKGEVPGGLAGSPGIPG
VKEQGFMPGPGQPGQPGTGHVPGKPGDQPGQPGQGLPGHPCPMGPGFPGGIN
GPKDQNGQWAPGVPKPGDGFQGMPTGSGPTQSGMDMLGVLPGVFGQKQK
LPGQGVKGDQGVPGPKLQGGPPGPGPYDVIKEPGLPGEPGPGKLGQGGPG
PKQQGVTSVGLPGPPGPGFPGDQKQKGTGPGPGPRGPGPPGPGDGLPGSMGP
PGTSPDHGFLVTRHSQTTDDPLCPGPKILYHGYSLLYVQGNRAHQDLGTAGSCL
RKFTSPDLFCNNVNCNPNFASRNDYSYMLSTPRPMPMSMAPISGDNTRPTISCAVCE
APAMVAVHSQTIQIPCCPNHSSSLWIGYSFVMTSAGAECSQAALASPCSLLEFPS
APTECHGRGTCTNYANAYSWFLATIERSEMFKNKPTSTLKAGELRTHVSRQVCMRR
T"
132..212
    /gene="Col4a-1"
213..5138
    /gene="Col4a-1"
    /product="alpha-1 type IV collagen"
5406..5411
    /gene="Col4a-1"
    /note="putative"
5732..5737
    /gene="Col4a-1"
    /note="putative"
5761..5766
    /gene="Col4a-1"
    /note="putative"
5765..5770
    /gene="Col4a-1"
    /note="putative"
5777..5782
    /gene="Col4a-1"
    /note="putative"
5858..5863
    /gene="Col4a-1"
    /note="putative"
6512
    /gene="Col4a-1"
    /note="putative"
BASE COUNT 1553 a 1798 c 1950 g 1211 t
ORIGIN
Query Match 35.4%; Score 36.8; DB 10; Length 6512;
Best Local Similarity 61.5%; Pred. No. 0.64;
Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
QY 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGCTCTCTGGTGCACAGGTT 61
Db 2692 CAGGACAGTCAGGCTCCCTGGCTTCTCTGGACAGGAGCACCTGGAGTTCAGGTT 2751
QY 62 TACCACAATATACAGAGAATAAGTGAATGACAA 97
Db 2752 TCCAGGTTCTAAAGGTGAATGGTGTCTATGGGAA 2787

RESULT 7
HNSCOL1
LOCUS HNSCOL1 694 bp mRNA linear INV 30-JUN-1993
DEFINITION Hydra N-COL 1 mRNA for mini-collagen.
ACCESSION X61045 S66747
VERSION X61045.1 GI:9446
KEYWORDS collagen homologue; N-COL 1 gene; nematocyst.
```

[illegible]

```
6453..6566))
/genes="F57B1.1"
/complement(join(5920..6069,6126..6209,6280..6408,
6453..6566))
/genes="F57B1.1"
/notes="predicted using Genefinder"
/codon_start=1
/protein_id="CAB01510.1"
/db_xref="GI:3877752"
/translation="MNNCOYEDERILMKHLYELILKQAISSVFLCQTIKPEIATFL
FILPILGIIYATIIKIILMKKALNKYELILKQAISSVFLCQTIKPEIATFL
IKRIINTTEICAGATPCPFRTTSKEIRKLVSRVSLTISQANSQQRRQTSRAI"
/complement(join(10251..10976,11025..11240))
/genes="F57B1.3"
/complement(join(10251..10976,11025..11240))
/genes="F57B1.3"
/notes="predicted using Genefinder"
/contains similarity to Pfam domain: PF01391 (Collagen
triple helix repeat (20 copies)), Score=45.4,
E-value=4e-10, N=3; PF01484 (Nematode cuticle collagen
N-terminal domain), Score=40.6, E-value=1.1e-08, N=1
cDNA EST yk73c7.5 comes from this gene"
/codon_start=1
/protein_id="CAB01509.1"
/db_xref="GI:3877751"
/translation="MSSEDPKQIAYETESLRKVPFGIAVSTIATLTAVVPMVLYN
MHOVSSLOSEVEFCANHSRGLWDEYOREGVSAGVAGHKRESYHRKARSKVRGSI
GADAAVGGSGAGSCSCGTGAGPAGSPQDQAGNDGAPGAPNGQDASEDOT
AGPDSFCFCPCAGPGSPGAGKPGSPGAPGPGGTGGASLPFGPPGAPGPPGPGQPG
SNGNAGAPQVDPVQPTGPGPAGPPGPPGAPGPPGQSGSGQPGGPGQGDAGA
PCAPGQPGAGQGDGDSGECADCHCPRPTAPCY"
/complement(join(13023..13760,13809..13948,14082..14274))
/genes="F57B1.4"
/complement(join(13023..13760,13809..13948,14082..14274))
/genes="F57B1.4"
/notes="predicted using Genefinder"
/contains similarity to Pfam domain: PF01391 (Collagen
triple helix repeat (20 copies)), Score=44.2,
E-value=9.7e-10, N=3; PF01484 (Nematode cuticle collagen
N-terminal domain), Score=40.6, E-value=1.1e-08, N=1
cDNA EST yk93c11.3 comes from this gene
cDNA EST yk70b10.3 comes from this gene
cDNA EST yk93c11.5 comes from this gene
cDNA EST yk102e6.3 comes from this gene
cDNA EST yk100f2.5 comes from this gene
cDNA EST yk39e7.3 comes from this gene
cDNA EST yk99e7.5 comes from this gene
cDNA EST yk72e11.3 comes from this gene
cDNA EST yk70b10.5 comes from this gene
cDNA EST yk102e6.5 comes from this gene
cDNA EST yk72e11.5 comes from this gene
cDNA EST yk170f7.5 comes from this gene
cDNA EST yk170f7.3 comes from this gene
cDNA EST yk279e6.3 comes from this gene
cDNA EST yk279e6.5 comes from this gene"
/codon_start=1
/protein_id="CAB01508.1"
/db_xref="GI:3877750"
/translation="MAETGVIDKAVEGGEKPGNSWFELASALPYVYLSEYVTLASI
KGMTQWSNHLTRSCPFHSFTIATLTAVVPMVLYNMOHVQSLQSEVEFCFAHRS
NGLWYORFEGVSGVAGRIKRESYHRASGPHRASKVRROSYGADAAVCGVGSAGG
SCSCGTGAGPAGSPGQDAGPCNDGAPGAPNGQDASEDTAGDFSCFCPCAGPP
GFSGAPGQKGSAGPAGPQSGASLPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
DVGTPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
DGSSEGCADCHCPRPTAPCY"
/complement(join(19688..19768,19816..19867,19913..20010,20397..20489,
20538..20687))
/genes="F57B1.5"
/complement(join(19688..19768,19816..19867,19913..20010,20397..20489,
20538..20687))

/genes="F57B1.5"
/notes="predicted using Genefinder"
/codon_start=1
/protein_id="CAB01512.1"
/db_xref="GI:3877754"
/db_xref="SPTREMBL:Q20925"
/translation="MPGENFKYFYVFPVFNVLISIKITIIILILSLYQWDTKORIS
GSNFLAIYDILAIGIASISLWFIQKVFYKVESFNIIGILSLASISFTITLD
PASIQSLTHILTLIIWISAFMAWRETYNNVSNLFISEWQGGFETIEYLEQHF"
/complement(join(23062..23112,24099..24151,24322..24499,24547..24696,
24748..24830,25045..25267))
/genes="F57B1.6"
/complement(join(23062..23112,24099..24151,24322..24499,24547..24696,
24748..24830,25045..25267))
/genes="F57B1.6"
/notes="predicted using Genefinder"
/cDNA EST EMBL:T01341 comes from this gene"
/codon_start=1
/protein_id="CAB01513.1"
/db_xref="GI:3877755"
/db_xref="SPTREMBL:Q20926"
/translation="MTMNVGGINLKNLYVIDAVQCKLORSFSAVTLHLCETRLGDSRL
YSRLIIDRKSEPIVYKHQIYVSIISGLTSSVFDFIVVSRLYYTPVDHKVLINPRM
AFWLFPSTFSLSLFSLFSLFSGAPRVIRKFAVDNQMLSLAHALMCGACILSSFVS
FICAQNANDIAGYAATPAQFQALASAWYYSRLRALVVISVIOSSLNGIIVSTLYLGI
SKYRTFVQLAERKPPQGLIERTTYFA"
/complement(join(34555..34692,34753..35192,35241..35370))
/genes="F57B1.7"
/complement(join(34555..34692,34753..35192,35241..35370))
/genes="F57B1.7"
/notes="predicted using Genefinder"
/codon_start=1
/protein_id="CAB01507.1"
/db_xref="GI:3877749"
/db_xref="SPTREMBL:Q20920"
/translation="MADQIKRRQRKQRLTLTAKNSITMTPSPNOISLNALNSLLY
GSINTSPQLLSPTSDASSCSMPFTPIPMFMTSADCSPTQPTQPTSPSSIA
STSPMTSLPLRLSGPLNLRDPSAEASLLNLCNADAALILTDQYRMILEEAM
SMSSSPSKDSDGSDGLNSIIVITDYPICLSRAFLVLSNYPRFHPPLHLTLD
CDYSAHCDLPCGPFLFS"
/... BASE COUNT 11789 a 6371 c 6437 g 11935 t
ORIGIN

Query Match 34.8%; Score 36.2; DB 3; Length 36532;
Best Local Similarity 65.4%; Pred. No. 1.2;
Matches 53; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 1 GCTGGACCTCCAGCACACCCAGCCGCCAGAGAAGTGGGGCCTCTGTGGTGCACCAAGT 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10679 GCTGGACCAACAGGCCCATCAGGAGCCAGGACCAAAAGGACCTTCAGGAGCCCAAGGA 10620

QY 61 TTACCAACATATACAGGAGAA 81
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10619 GTCCAGGACAAACCCGGAGGA 10599

RESULT 9
CEY51H4A/c
LOCUS CEY51H4A 269619 bp DNA linear INV 26-APR-2002
DEFINITION Caenorhabditis elegans cosmid Y51H4A, complete sequence.
ACCESSION AL132952 AL031823
VERSION AL132952.1 GI:6434491
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1
REFERENCE
AUTHORS none.
TITLE Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 99069613
```

PUBMED 9851916  
 REMARK The C.elegans Sequencing Consortium.  
 REFERENCE 2 (bases 1 to 269619)  
 AUTHORS Sulston J.E.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-OCT-1999) Nematode Sequencing Project, Sanger  
 Institute, Hinxton, Cambridge CB10 1SA, England and Department of  
 Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:  
 jes@sanger.ac.uk or rv@nematode.wustl.edu  
 On May 14, 2001 this sequence version replaced gi:5730145.  
 COMMENT Coding sequences below are predicted from computer analysis, using  
 predictions from Genefinder (P. Green, U. Washington), and other  
 available information.  
 Current sequence finishing criteria for the C. elegans genome  
 sequencing consortium are that all bases are either sequenced  
 unambiguously on both strands, or on a single strand with both a  
 dye primer and dye terminator reaction, from distinct subclones.  
 Exceptions are indicated by an explicit note.  
 For a graphical representation of this sequence and its analysis  
 see:- [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=y51H4A)  
 name=y51H4A  
 IMPORTANT: This sequence is NOT necessarily the entire insert of  
 the specified clone. It may be shorter because we only sequence  
 overlapping sections once, or longer because we arrange for a small  
 overlap between neighbouring submissions.  
 This sequence is the entire insert of clone y51H4A. The true left  
 end of clone y43D4A is at 193825 in this sequence. The start of  
 this sequence (1..115) overlaps with the end of sequence AL713992.  
 The end of this sequence (269514..269619) overlaps with the start  
 of sequence AL132846.  
 FEATURES  
 source Location/Qualifiers  
 1. .269619  
 /organism="Caenorhabditis elegans"  
 /db\_xref="taxon:6239"  
 /chromosome="IV"  
 /clone="Y51H4A"  
 complement(join(2293..2346,2393..2610,2668..2758,  
 3398..3694,5250..5357))  
 /gene="Y51H4A.1"  
 complement(join(2293..2346,2393..2610,2668..2758,  
 3398..3694,5250..5357))  
 /genes="Y51H4A.1"  
 /note="predicted using Genefinder"  
 /codon\_start=1  
 /protein\_id="CAB61134.1"  
 /db\_xref="GI:6434492"  
 /db\_xref="SPTREMBL:Q9NAE5"  
 /translation="MLLVLLIFTLLFFYFGKLIIFGKIVENSIEKINSNGIS  
 LAENSKNLEKSSNDPENAEETEMVIELHTEAPVNYLEEEIQIKVPEVEDETEK  
 PAPSFKSTSSCISEILSEQSIGTCARLYFTNRPLYGFRHTAEPVCCVLARFQFO  
 NFNCFKFSISPRSPSPPLDFATSSYAPLSNFTRETPNIVPRMSEFEFLHKCIEK  
 RKSDIRDLRIQAKRRHRFRSEHLQOVQDHIDY"  
 complement(join(12157..12228,13962..14312,16459..16593,  
 17111..17131))  
 /gene="Y51H4A.3"  
 complement(join(12157..12228,13962..14312,16459..16593,  
 17111..17131))  
 /genes="Y51H4A.3"  
 /note="contains similarity to Pfam domain: PF000071 (Ras  
 family), Score=317.2, E-value=6.4e-92, N=1  
 cDNA EST yk60d10.5 comes from this gene  
 cDNA EST yk103g7.5 comes from this gene  
 cDNA EST yk77f3.5 comes from this gene  
 cDNA EST yk236a7.5 comes from this gene  
 cDNA EST yk435f7.5 comes from this gene  
 cDNA EST yk531a12.5 comes from this gene  
 cDNA EST yk592g8.5 comes from this gene  
 cDNA EST yk651f5.5 comes from this gene  
 cDNA EST yk671f3.5 comes from this gene  
 cDNA EST yk565h1.5 comes from this gene"  
 /codon\_start=1  
 /protein\_id="CAB63379.1"  
 /db\_xref="GI:6580310"  
 /translation="MAAIRKKLVIVGDGACGKCTCLLIIVFSKDDQPDVVYVTFVFNVA  
 DIEVDGQVELALWDTAGEDYDRRLPSYPTDVIILMCFSIDSPDLSLENPEKWTPE  
 VRHECPNVPVILLYGNKKDLRSDPOTVRELAKMKQEPVKPEQGRAIAEQIGAFVLECS  
 AKYDGRVFERATQAALQOKKKKSKCMILL"  
 complement(join(21593..21756,22186..22384,22466..22675))  
 /gene="Y51H4A.2"  
 complement(join(21593..21756,22186..22384,22466..22675))  
 /genes="Y51H4A.2"  
 /codon\_start=1  
 /protein\_id="CAB63380.1"  
 /db\_xref="GI:6580311"  
 /db\_xref="SPTREMBL:Q9U265"  
 /translation="MKSVGGGSRLLLLLSRRRHSISLLPLLGGRFPSLSHFDPHIHY  
 IHGCLAAPREPQOKGALDFKFOHSHATATLCCGRVPLLYARKLSEPHWGHARRSS  
 VDNGKKRRRGKAKPIILTOFSKSFLEFLDKMYKNEVELSYFKSSIFCTFFPKKITPFR  
 LLDGIPAPLKLVSRAAERERERAAITRIV"  
 complement(join(31618..31755,32295..32402,33420..33614,  
 33718..33834))  
 /gene="Y51H4A.4"  
 complement(join(31618..31755,32295..32402,33420..33614,  
 33718..33834))  
 /genes="Y51H4A.4"  
 /note="predicted using Genefinder"  
 cDNA EST yk282h7.3 comes from this gene  
 cDNA EST yk340d3.3 comes from this gene  
 cDNA EST yk479g1.3 comes from this gene  
 cDNA EST yk282h7.5 comes from this gene  
 cDNA EST yk340d3.5 comes from this gene  
 cDNA EST yk479g1.5 comes from this gene  
 cDNA EST yk486f5.5 comes from this gene  
 cDNA EST yk573a7.3 comes from this gene  
 cDNA EST yk573a7.5 comes from this gene"  
 /codon\_start=1  
 /protein\_id="CAB61136.1"  
 /db\_xref="GI:6434494"  
 /db\_xref="SPTREMBL:Q9NAE4"  
 /translation="MMTDIOQLAASRMCDSDSIKAFNQMYAQNYPYAGAGFDKNYSAA  
 TALPFFAQHEAPYANPHATSFSPSSSTSTTSSONQSGSSKKKKPVVPENOK  
 DEAYLDRRRRNNEAARKSRERKKVQDQNSVRVYLERENQCLRVYVQLOLQNESMR  
 QHLLQNPGTIDTMDPIAMANMPSY"  
 complement(join(50748..50898,51588..51977,52741..52920,  
 53245..53366,53557..53611,54196..54274,54700..54854,  
 55009..55055,57258..57575))  
 /gene="Y51H4A.5"  
 complement(join(50748..50898,51588..51977,52741..52920,  
 53245..53366,53557..53611,54196..54274,54700..54854,  
 55009..55055,57258..57575))  
 /genes="Y51H4A.5"  
 /note="predicted using Genefinder"  
 contains similarity to pfam domain: PF01764 (Lipase),  
 Score=244.3, E-value=5.4e-70, N=1"  
 /codon\_start=1  
 /protein\_id="CAB61137.1"  
 /db\_xref="GI:6434495"  
 /db\_xref="SPTREMBL:Q9NAE3"  
 /translation="MERLNCOLYLLHLLFFLYVCCWIVGTSGALGRAEYTCHEKNTINL  
 DKDATREMSDHLKMYRTYQWCDNTSLKYSRFOILHSNALRIETSDAVGSAIILSKL  
 LFDVKNNTIKKKHNFKQKKYRTFENVEFKFREKLINLELFFACFNTCLFRNRIIF  
 FNSAANLNFATILVFLIFWQFFDLEAKSCDTCIAGKKHCWCVTNBEGGSPCATSIS  
 KSLNCPGLPKAHAYDDNFRTKIVPLAASALNPENPEKCFKPRLPTLTKKFSANCS  
 EVGQSCFSTSPDTTKQVLVMSFRGTDSPLQTLDEITLDFFTGKGQFPDPAGNIFTY  
 FYDAFFLWAGLQDDIROLKYKVPDYELWVTHSLGALASVAASYVHTGIFTSDK  
 KLVTLQCPRTGDDYATWHDKNFPYSFRIVHRDLVAHIPQDQADKLFHHRSEVWY  
 NNNMSTPEPHICAEADGFCSSRELDNTNADHLTYFGVIFPGWQTKC"  
 join(61826..61986,62058..62127))  
 /gene="Y51H4A.6"  
 join(61826..61986,62058..62127))  
 /genes="Y51H4A.6"  
 /codon\_start=1  
 /protein\_id="CAB63381.1"  
 /db\_xref="GI:6580312"  
 /db\_xref="SPTREMBL:Q9U264"





[illegible]



Db 17444 GCTGATCCCGAGGACCGGACATCCAGGAACCGCGGAGCTCCAGGACCCAGGA 17385

Qy 61 TTACCACAATATACAGGAG 79

Db 17384 GCACCGAGAAATCCAGGAG 17366

## RESULT 15

HSC3A1R  
LOCUS Human COL3A1 mRNA for pro alpha-1 (III) collagen. linear PRI 05-AUG-1995  
DEFINITION  
ACCESSION X15332  
VERSION X15332.1 GI:29545  
KEYWORDS COL3A1 gene; collagen.  
SOURCE Homo sapiens.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 3234)  
Janeczko.R. and Ramirez,F.  
Direct Submission  
TITLE Submitted (19-MAY-1989) Janeczko R., Ramirez F., Suny Health  
JOURNAL Science Centre, 450 Clarkson Avenue - Box 44, Brooklyn NY 11203, U  
S A

REFERENCE 2 (bases 1 to 3234)  
AUTHORS Janeczko.R.A. and Ramirez,F.  
TITLE Nucleotide and amino acid sequences of the entire human alpha 1  
(III) collagen  
JOURNAL Nucleic Acids Res. 17 (16), 6742 (1989)  
MEDLINE 89386015  
PUBMED 2780304

COMMENT The sequence overlaps with that reported by Chu et. al. in J. Biol.  
Chem. 260:4357-4363(1985), by Toman et. al. in Nucl. Acids Res.  
16:7201-7201(1988) and by Mankoo et. al. in Nucl. Acids Res.  
16:2337-2337(1988).

## FEATURES

## source

1..3234  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="2q31"  
<1..>3234  
/codon\_start=1  
/product="alpha-1 (III) collagen"  
/protein\_id="CAA33387.1"  
/db\_xref="GI:930045"  
/db\_xref="SWISS-PROT:P02461"  
/translation="QNYSPQYDSYVKSQVAVYGLAGYCPAGPGPPGPPGTSGHP

## CDS

GSPPGTFQPPGEPGAGSPGPPGPAIGSPGACKGDSGRPRGDRGLPGPPG  
IKGPAGIPGPGMKHGRGKEGTAGLKGNGLPGENGAPGPMGPRGAPGE  
RGRFGLPAAAGARGNDGARGSDGPPGPPGTAGFPSPGAKGEVGPAGSPGNSGAP  
GQREPPQGHAGAOGPPGPPGPGINGSPGKGMGPAGIPGAPGLMGARGPPGPAGANG  
APGLRGAGEPGKNGAKGEPGPRGERCEAGIPGVPGAKGEDKDGSPGDPGANGLPGA  
AGRGALGSRGAPAGNIPGKGPAGERGAPGAPGACGAEGFRDGPVPGPMRMP  
GSPPGQSKGPPGSGESGRPPGPPGSPGPGVGMGPPGPGNDGAPGNRNG  
PGPGPGPPGKNGEYGPQPGTGGDKGTGPRGGLGQLGPGTGGPGENGK  
PGFPKRGAGACGAGKGDAGAPGERGPPGLAGPLRGAGPPGEGGKGAAGPP  
PGPAAGTGLQGMPEGRGLSPGPKDGEPGPGDGVPGKDGPRGTPGIPGPG  
PAGPGDKGEGAGPLPIAGPRGSPGERGETGPPGAPGAPGQNGEPGKGERCA  
PGKGECCGPPGAVGPGGSPAGPPGPGQVKGERSPGGPGCAAGFPGARLPGPPGSN  
GNPFGPGSPGKRDGPPGAGNTGAPSPGVSGPGDAGQPGKEKSPGAOGPPGAPG  
PLGIAGTTGARGLAGPMPGPRGSPGVKGESKPGANGLSGERGPPGQGLPL  
AGTAGEPGRDGNPGSDGLPDGSPGKDRGENSGPAPGAPGHPGPPGVPGAPGS  
GDGSEGPAGAPGAPGAGSKGAPGPGPRGDKGETGERGAAGIKGHRGPPGNPAGPG  
SPGAPGQGAIGSPGAPGPRGVPVGPSPGCKDGTSGHPGPIGPRGPRNGRGERGSEGS  
PGHPGSGPPGPGAGPCGCGGCAPAIAGIAGKAGGAFPIYGDEPM"

664 a 861 c 1106 g 603 t

BASE COUNT  
ORIGIN

Query Match 32.7%; Score 34; DB 9; Length 3234;

Best Local Similarity 61.1%; Pred. No. 4.5; Mismatches 0; Gaps 0;  
Matches 55; Conservative 0; Indels 35; Gaps 0;

Qy 1 GCTGGACCTCCAGCACACCCCGAGGCCCCAGAGAAGTGGGGCTCTGCTGTCACCCAGGT 60

Db 85 GCTGCCCCCCCCAGGCCCCCCCCCGCCCCCTGCTACATCTGTCATCTGTTCCCTGGA 144  
Qy 61 TTACCACAATATACAGGAGAGAAATAAGTGAA 90  
Db 145 TCTCCAGGATACCAAGGACCCCTGGTGAA 174

Search completed: February 20, 2003, 04:31:48  
Job time : 500.581 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:48:42 ; Search time 1130.07 Seconds  
(without alignments)  
6406.126 Million cell updates/sec

Title: US-09-997-610-1\_COPY\_2\_448

Perfect score: 447

Sequence: 1 atagtggtacatactgtctt.....cttggtgtctttaaggca 447

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estlin:\*
  - 4: em\_estmu:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_htc:\*
  - 9: gb\_est1:\*
  - 10: gb\_est2:\*
  - 11: gb\_htc:\*
  - 12: gb\_est3:\*
  - 13: gb\_est4:\*
  - 14: gb\_est5:\*
  - 15: em\_estfun:\*
  - 16: em\_estom:\*
  - 17: gb\_gss:\*
  - 18: em\_gss\_hum:\*
  - 19: em\_gss\_inv:\*
  - 20: em\_gss\_pln:\*
  - 21: em\_gss\_vrt:\*
  - 22: em\_gss\_fun:\*
  - 23: em\_gss\_mam:\*
  - 24: em\_gss\_mus:\*
  - 25: em\_gss\_other:\*
  - 26: em\_gss\_pro:\*
  - 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	113.6	25.4	499	14	BM967732
2	76.4	17.1	504	12	BE757275
3	76.4	17.1	526	12	BE757276
4	61.8	13.8	612	17	AG088117
5	59.6	13.3	538	9	AA777621
6	58.6	13.1	440	10	BE062167

c 7	58.6	13.1	475	17	B36584
c 8	58.6	13.1	537	10	BE079777
c 9	58.6	13.1	563	10	BE079876
c 10	58.6	13.1	678	10	AW813783
c 11	58.6	13.1	727	10	AV731140
c 12	58.6	13.1	2615	10	BE420422
c 13	58	13.0	342	17	AQ012504
c 14	58	13.0	345	12	BF881529
c 15	58	13.0	670	17	AG093990
c 16	57.8	12.9	397	13	BI180953
c 17	57	12.8	425	10	AW517269
c 18	57	12.8	426	9	AA372508
c 19	57	12.8	660	17	AG019007
c 20	57	12.8	666	17	AG141320
c 21	57	12.8	683	17	AG019080
c 22	57	12.8	1036	13	BM462674
c 23	56.8	12.7	577	10	AV716883
c 24	56.6	12.7	662	10	BE390081
c 25	56.4	12.6	284	9	AA360873
c 26	56.4	12.6	461	12	BF828364
c 27	56.4	12.6	461	12	BF829000
c 28	56.4	12.6	545	10	BE279213
c 29	56.4	12.6	609	10	BE389158
c 30	56.4	12.6	667	17	AG080755
c 31	56.4	12.6	668	17	AG080765
c 32	56.4	12.6	698	12	BE729966
c 33	56.4	12.6	705	17	AG052122
c 34	56.4	12.6	721	12	BE728927
c 35	56.4	12.6	761	12	BE728912
c 36	56.4	12.6	795	12	BF025976
c 37	56.4	12.6	800	12	BF203746
c 38	56.4	12.6	885	12	BF304683
c 39	56.4	12.6	1080	14	BQ422247
c 40	56	12.5	315	9	AA719635
c 41	56	12.5	645	17	AG014562
c 42	55.8	12.5	360	10	AW898344
c 43	55.4	12.4	364	17	AQ665181
c 44	55.4	12.4	388	17	B78843
c 45	55.4	12.4	453	17	AQ393565

ALIGNMENTS

RESULT 1	BM967732	499 bp	mrna	linear	EST 20-MAR-2002
LOCUS	LM24HW0134	Bos taurus	LM-24-HW	cdna	library Bos taurus cdna clone
DEFINITION	LM-24-HW-011-34 (5'), mRNA sequence.				
ACCESSION	BM967732				
VERSION	BM967732.1				GI:19561919
KEYWORDS	EST.				
SOURCE	COW.				
ORGANISM	Bos taurus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.				
AUTHORS	Yoon,D.H., Jang,Y.S., Kim,T.H., Park,E.W., Lee,H.K., Chung,E.R., Sun,S.S. and Cheong,I.C.				
TITLE	Gene Expression Profiling of the Bovine skeletal muscle				
JOURNAL	Unpublished (2002)				
COMMENT	Contact: Dr. Du-Hak Yoon National Livestock Research Institute, RDA 564 Omoekchun-dong, Suwon, 441-350, Korea Tel: 82 31 290 1593 Fax: 82 31 290 1792 Email: dhyoon@rda.go.kr Insert Length: 499 Std Error: 0.00 Seq primer: CAGGAACACGCTATGAC POLYA-No.				
FEATURES	Location/Qualifiers				
source	1. .499				

```
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="LM-24-HW-011-34 (5')"
/clone_lib="Bos taurus LM-24-HW cDNA library"
/sex="six males mixed"
/tissue_type="myocyte"
/dev_stage="24 months old"
/lab_host="XLI-BlueMRF'strain"
/note="Organ: skeletal muscle; Vector: Uni-ZAPXR; Site_1:
EcoRI; Site_2: Xho I"
BASE COUNT 129 a 139 c 131 g 100 t
ORIGIN
Query Match 25.4%; Score 113.6; DB 14; Length 499;
Best Local Similarity 64.2%; Pred. No. 4.4e-21;
Matches 203; Conservative 0; Mismatches 109; Indels 4; Gaps 2;
QY 46 GAAGTTGCTGGACCTCCAGCACACCCAGGCCCCAGAGAGTGGGGCTCTCTGGTGCA 105
Db 174 GGACCTCCGGGCCACAGGACACCCAGGCCCTCCAGGTATTAAGAGGGCCCTCCAGGTATA 233
QY 106 CCAGGTTTACCATATATACAGGAG---AAATAAGTGAATGACAAAATGCCCTGTCT 162
Db 234 AGAGGAATACCAAGTTTGCCAGGTCCGCCGGGAAGTCCAGGACCAAGTGTAAATGCCCA 293
QY 163 GATATAGAAAGTGCAGCTTTACTGTGAAGCTCAGTGGAAACTTCCCTTCTTCAAG 222
Db 294 TCCACAGACAGTCGCCCTTCTACTGTGAAGCTCAGTGGCCAGTTGCCCTTCCCTTCAAAG 353
QY 223 CCCAT-CATCTTCACAGGGGCTCTGTACAATGCCAGAGGGATTTAAAGGAGGCCATGGG 281
Db 354 CTGTGCGCTTTCACAGAGGCTCTGTACAATGCCAGAGAGACTTACAGGAGGACACTGG 413
QY 282 AGCTTTGCTTCAGGGTGCCTGGGAATTAATCTACTCCAGCTTTGATGTAGCTGCATCA 341
Db 414 GCTCTTCATCAGGGTGCCAGGAATTAACCATTTCTCTTATGTGGATCTCCATCA 473
QY 342 TTGCAAGGTGATATT 357
Db 474 CTGCAAGGTGACTGTT 489
```

```
RESULT 2
BE757275
LOCUS BE757275 504 bp mRNA linear EST 25-APR-2001
DEFINITION 211668 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE757275
VERSION BE757275.1 GI:10171267
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 504)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keefe,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
```

```
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCCACTCAGGAGC
Plate: 63 row: G column: 13
Seq primer: ATTTAGGTGACACTATAG.
FEATURES
Location/Qualifiers
1..504
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 2BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from testis, thymus,
semitendinosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."
BASE COUNT 99 a 131 c 146 g 128 t
ORIGIN
Query Match 17.1%; Score 76.4; DB 12; Length 504;
Best Local Similarity 61.6%; Pred. No. 1.2e-10;
Matches 122; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
QY 174 GTCAGCCTTTACTGTGAAGCTCAGTGGAAACCTTCTCTTCTTCAAGCCCATCATCTT 233
Db 303 GTCTGCCCTTTCCGTCAAGCTGAGTGGGCTTTCCAGGACCTCCAGCCCATTTGCTT 362
QY 234 CACAGGGTCTGTACAATGCCAGAGGGATTTAAAGGAGGCCATGGGAGTCTTTGCTTG 293
Db 363 CCAGGAAGTCTGTACAACCATCAGGGCCACTTCGACCCGCCACTGGTGTCTTCAAGTG 422
QY 294 CAGGCTGCTGGGAATTACTACTCCAGCTTTGATGTAGCTGCATCATTCGAAGTGAA 353
Db 423 CAGCGTCTGCTGTGTACCACTTTGGCTTTGACATTTGAGTTGTTTCAGAGTGTCTCAA 482
QY 354 TATTTGGCTAATGAGGAA 371
Db 483 GGTGGTCTAATGCGGAA 500
RESULT 3
BE757276
LOCUS BE757276 526 bp mRNA linear EST 25-APR-2001
DEFINITION 211669 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE757276
VERSION BE757276.1 GI:10171268
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 526)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keefe,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
PCR Primers
```









```

Db 553 AAGCATTCAGAGGTAACCTTGGGTGCTGTTAAAGGCA 589
RESULT 13
AQ012504/c
LOCUS
DEFINITION CIT-HSP-2298M4.TF CIT-HSP Homo sapiens genomic clone 2298M4, DNA
sequence.
ACCESSION AQ012504
VERSION AQ012504.1 GI:3185069
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 342)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL Unpublished (1998)
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
FEATURES
source
1..342
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GB:7153856"
/db_xref="taxon:9606"
/clone="2298M4"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelosBAC11; Site_1: HindIII; Site_2:
HindIII"
BASE COUNT 88 a 85 c 58 g 111 t
ORIGIN
Query Match 13.0%; Score 58; DB 17; Length 342;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 390 GGAAGAAATTTCTAAGCAGCAAGCATTCAGAGGTGCTGCTGTTAAAGGCA 447
|||||
Db 264 GGAAGAAATTTCTAAGCAGCAAGCATTCAGAGGTGCTGCTGTTAAAGGCA 207
|||||
RESULT 14
BF881529/c
LOCUS
DEFINITION QV1-ET0181-031200-546-f03 ET0181 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF881529
VERSION BF881529.1 GI:12271655
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 345)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV1et2-QV1-ET0181-
031200-546-f03et3-2000-12-03et4-1)
Seq primer: puc 18 forward
High quality sequence stop: 89.
FEATURES
source
1..345
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ET0181"
/dev_stage="Adult"
/note="Organ: lung_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No.196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 86 a 96 c 57 g 106 t
ORIGIN
Query Match 13.0%; Score 58; DB 12; Length 345;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 390 GGAAGAAATTTCTAAGCAGCAAGCATTCAGAGGTGCTGCTGTTAAAGGCA 447
|||||
Db 194 GGAAGAAATTTCTAAGCAGCAAGCATTCAGAGGTGCTGCTGTTAAAGGCA 137
|||||
RESULT 15
AG093990/c
LOCUS
DEFINITION Pan troglodytes DNA, clone: PTB-094J23.R, genomic survey sequence.
ACCESSION AG093990
VERSION AG093990.1 GI:16645792
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC Library clone:PTB-094J23.R.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 670)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chuhmpbes@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of

```

clone tracking errors.

PRIMERS

Sequencing: M13Rev

LIBRARY

Vector : PKS145

R.Site 1 : SacI

R.Site 2 : SacI.

Location/Qualifiers

1. .670

/organism="Pan troglodytes"

/db\_xref="taxon:9598"

/clone="PTB-094J23.R"

/sex="male"

/cell\_type="lymphoblast"

/clone\_lib="PTB Chimpanzee Male BAC Library"

155 a 186 c 124 g 205 t

ORIGIN

BASE COUNT

Query Match

Best Local Similarity 13.0%; Score 58; DB 17; Length 670;

Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 390 GGAAGAAATTTCTAAGCAGCAAGCATTCAAGAGGTGACTTGGGTGCTGTTAAAGGCA 447

Db 404 GGAGAAATTTCTAAGCAGCAAGCATTCAAGAGGTGACTTGGGTGCTGTTAAAGGCA 347

Search completed: February 20, 2003, 06:14:39

Job time : 1136.07 secs



```
|||||
Db 182 TTCTACGTGAAGCTCAGTGAAAACCTCTCTCTCCCTTTCAAGGCCATCATCTTCACAGGG 241
QY 241 GTCCTGTACAATGCCAGAGGGATTTAAAGAGGCCATGGAGCTCTTGTCTTCAGGGTG 300
Db 242 GTCCTGTACAATGCCAGAGGGATTTAAAGAGGCCATGGAGCTCTTGTCTTCAGGGTG 301
QY 301 CTTGGGAATTTACTACTCCAGCTTTGTAGTTGAGCTGCATCATCTGCAAGTGAATATTTGG 360
Db 302 CTTGGGAATTTACTACTCCAGCTTTGTAGTTGAGCTGCATCATCTGCAAGTGAATATTTGG 361
QY 361 CTAATGAGGAAGCAAAATTTGGCTAATAAGGAAGAAATTTCTAAGCAGCAAAAGCATTTCAA 420
Db 362 CTAATGAGGAAGCAAAATTTGGCTAATAAGGAAGAAATTTCTAAGCAGCAAAAGCATTTCAA 421
QY 421 GAGGTGACTTGGGTGCTGTTAAAGGCA 447
Db 422 GAGGTGACTTGGGTGCTGTTAAAGGCA 448

RESULT 2
US-09-997-610-5
; Sequence 5, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 1731
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1731)
US-09-997-610-5

Query Match 100.0%; Score 447; DB 9; Length 1731;
Best Local Similarity 100.0%; Pred. No. 3.5e-130; Mismatches 0; Indels 0; Gaps 0;
Matches 447; Conservative 0;

QY 1 ATAGTGGTCATACCTGCTCTTAATAACGGCAGTCATTTGAGCATGTAGAAGTTGCTGGACCT 60
Db 355 ATAGTGGTCATACCTGCTCTTAATAACGGCAGTCATTTGAGCATGTAGAAGTTGCTGGACCT 414
QY 61 CCAGCACACCCAGGCCCCAGAGAAGTGGGCCCTCTGCTGCACACAGTTTACCACAA 120
Db 415 CCAGCACACCCAGGCCCCAGAGAAGTGGGCCCTCTGCTGCACACAGTTTACCACAA 474
QY 121 TATACAGGAGAAATAAGTGAATGACAAAATGCCCTGCTCTATATAGAAAGGTGACGC 180
Db 475 TATACAGGAGAAATAAGTGAATGACAAAATGCCCTGCTCTATATAGAAAGGTGACGC 534
QY 181 TTCTACTGTGAAGTCAAGTGGAAAACCTCTCTCTCTTTTCAAGCCCATCATCTTTCACAGGG 240
Db 535 TTCTACTGTGAAGTCAAGTGGAAAACCTCTCTCTCTTTTCAAGCCCATCATCTTTCACAGGG 594
QY 241 GTCCTGTACAATGCCAGAGGGATTTAAAGAGGCCATGGAGCTCTTGTCTTCAGGGTG 300
Db 595 GTCCTGTACAATGCCAGAGGGATTTAAAGAGGCCATGGAGCTCTTGTCTTCAGGGTG 654
QY 301 CTTGGGAATTTACTACTCCAGCTTTGTAGTTGAGCTGCATCATCTGCAAGTGAATATTTGG 360
Db 655 CTTGGGAATTTACTACTCCAGCTTTGTAGTTGAGCTGCATCATCTGCAAGTGAATATTTGG 714
```

```
QY 361 CTAATGAGGAAGCAAAATTTGGCTAATAAGGAAGAAATTTCTAAGCAGCAAAAGCATTTCAA 420
Db 715 CTAATGAGGAAGCAAAATTTGGCTAATAAGGAAGAAATTTCTAAGCAGCAAAAGCATTTCAA 774
QY 421 GAGGTGACTTGGGTGCTGTTAAAGGCA 447
Db 775 GAGGTGACTTGGGTGCTGTTAAAGGCA 801

RESULT 3
US-09-997-610-3
; Sequence 3, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1377
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide encoding a polypeptide
; OTHER INFORMATION: of SEQ ID NO:2
; NAME/KEY: variation
; LOCATION: (1)...(1377)
; OTHER INFORMATION: Each n is independently A, T, G, or C.
; NAME/KEY: misc_feature
; LOCATION: (1)...(1377)
; OTHER INFORMATION: n = A,T,C or G
US-09-997-610-3

Query Match 72.5%; Score 324.2; DB 9; Length 1377;
Best Local Similarity 61.0%; Pred. No. 1e-91;
Matches 272; Conservative 95; Mismatches 79; Indels 0; Gaps 0;

QY 1 ATAGTGGTCATACCTGCTCTTAATAACGGCAGTCATTTGAGCATGTAGAAGTTGCTGGACCT 60
Db 1 ATHGTGNTATHCCNGTNTYNATHACNGTNTATGARGAYTNGARGTNGNGNCCN 60
QY 61 CCAGCACACCCAGGCCCCAGAGAAGTGGGCCCTCTGCTGCACACAGTTTACCACAA 120
Db 61 CCNGCNCAYCCNMGNCNCCNGARGARGTNGNCCNCCNGNCCNGNNTNCCNCA 120
QY 121 TATACAGGAGAAATAAGTGAATGACAAAATGCCCTGCTCTATATAGAAAGGTGACGC 180
Db 121 TAYACNGNGARATHWSNGARATGACNAARTGYCCNTGYCCNGAYATHGARMNWSNGCN 180
QY 181 TTCTACTGTGAAGTCAAGTGGAAAACCTCTCTCTCTTTTCAAGCCCATCATCTTTCACAGGG 240
Db 181 TTCTACTGTGAAGTCAAGTGGAAAACCTCTCTCTCTTTTCAAGCCCATCATCTTTCACAGGG 240
QY 241 GTCCTGTACAATGCCAGAGGGATTTAAAGAGGCCATGGAGCTCTTGTCTTCAGGGTG 300
Db 241 GTNYTNTAAYCNCARMNGNAYYTNAARGARGCNATGGGNGTNTYGCNTGYMNGTN 300
QY 301 CTTGGGAATTTACTACTCCAGCTTTGTAGTTGAGCTGCATCATCTGCAAGTGAATATTTGG 360
Db 301 CCNGGNAATYATYAYWSNWSNTTYGAYTNGARYTNCAYCAYTGYAARGTNAAYATHHTGG 360
QY 361 CTAATGAGGAAGCAAAATTTGGCTAATAAGGAAGAAATTTCTAAGCAGCAAAAGCATTTCAA 420
Db 361 YTNATGMAARACARATHYTNNGCNAAAYARGARGARATHWSNARCARCARNATHCAR 420
QY 421 GAGGTGACTTGGGTGCTGTTAAAGGC 446
```



```
QY 163 GATATAGAAAGTCAACCTTACTGTGAAGCTCAGTGAAGAACTTCTCTCTCTTCAAG 222
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 276 TGCCACAGACAGTCGCCCTTCACTGTGAAGCTCAGTGGCCAGTTGCTTCCCTTCAAG 335
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 223 CCCATCATCTTCACAGGGTCTGTACATGCCACAGAGGATTTAAAGGAGGCCATGGGA 282
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 336 CCTGTGCCCTTCACAGAGTCTGTACAAATGCCAGAGACTTACAGGAGGACACTGGG 395
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 283 GTCTTTGCTTGCAGGGTGC 302
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 396 GTCTTCATGCAAGGTGC 415
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
US-09-960-352-220
; Sequence 220, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 220
; LENGTH: 389
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (277),(335)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 01-LIB34-084-Q1-E1-A9
US-09-960-352-220

Query Match      20.5%; Score 91.6; DB 10; Length 389;
Best Local Similarity 76.7%; Pred. No. 4.8e-19;
Matches 112; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 234 CACAGGGTCTCTACAATGCCAGAGGATTTAAAGGAGGCCATGGAGTCTTTGCTTG 293
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 CCCACGGTCCAGTACAAATGCCAGAGACATTACAGGAGGACTGGGCTCTTCACATG 60
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 294 CAGGGTGCCTGGGAATTACTTACTCCAGCTTTGATGTGAGTGCATCATTTGCAAGGTGAA 353
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 CAGGGTCCAGGAAATTAACCATTTCTCTTGATGTGGATCTCCATCACTGCAAGGTGAC 120
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 354 TATTTGGCTAATCAGGAAGCAATTT 379
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 TGTTCAGCTGATGAGGACAAAAGTT 146
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
US-09-960-352-1786
; Sequence 1786, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 1786
; LENGTH: 392
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (378)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 16-LIB34-049-Q1-E1-D11
US-09-960-352-1786

Query Match      17.9%; Score 80.2; DB 10; Length 410;
Best Local Similarity 54.6%; Pred. No. 1.9e-15;
Matches 185; Conservative 0; Mismatches 148; Indels 6; Gaps 1;
```

```
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 08-LIB34-023-Q1-E1-B7
US-09-960-352-1786

Query Match      18.5%; Score 82.8; DB 10; Length 392;
Best Local Similarity 61.4%; Pred. No. 2.8e-16;
Matches 151; Conservative 0; Mismatches 92; Indels 3; Gaps 1;

QY 46 GAAGTTGCTGGACCTCCAGCACACACCCAGCCCGCCAGAGAAAGTGGGGCTCTCTGTGTGCA 105
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 146 GGACCTCCGGGGCCACCAGGACACCCAGGGCTCCAGGTATAGAGGGCTCCAGGTATA 205
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 106 CCAGGTTTACCACAATATACAGGAG---AAATAAGTGAATGACAAAATGCCCTCTGCTCT 162
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 206 AGAGGAATACAGGTTTGCAGGTCGCCGGGNACTCCAGGACCAAGTGTAAATGCCCA 265
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 163 GATATAGAAAGTCAAGCCTTTACTGTGAAGCTCAGTGAAGAACTTCTCTCTCTTTCAAG 222
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 266 TGCCACAGACAGTCCGCCCTTCACTGTGAAGCTCAGTGGCCAGTTGCTTCCCTTCAAAG 325
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 223 CCCATCATCTTCACAGGGTCTGTACATGCCACAGAGGATTTAAAGGAGGCCATGGGA 282
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 326 CCTGTGCCCTTCACAGAGTCTGTACAAATGCCAGAGACTTACAGGATGACACTGGG 385
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 283 GTCTTT 288
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 386 GTCTTT 391
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
US-09-960-352-3715
; Sequence 3715, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 3715
; LENGTH: 410
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (378)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 16-LIB34-049-Q1-E1-D11
US-09-960-352-3715

Query Match      17.9%; Score 80.2; DB 10; Length 410;
Best Local Similarity 54.6%; Pred. No. 1.9e-15;
Matches 185; Conservative 0; Mismatches 148; Indels 6; Gaps 1;
```

```
Db 179 CAACCATGTCTTCAAGGAAGCCCTGCATAATGACCGACCACCTTCAATCTTACCACG 238
Qy 280 GGAGTCTTTGCTTGCAGGTCCTCGGAATTAATACTTCCAGCTTTGATCTTGAGCTGCAT 339
Db 239 GGAGTCTTCACTGCACCATCCTCGCTGCTGACCGCTTTGGCTTGACATAGAGCTATTC 298
Qy 340 CATTGCAAGGTGAATATTTGGCTAATGAGGAAGCAAAAT 378
Db 299 CAGCATGTGTGAAGCTAGGCTCATGAAGAATGACACT 337

RESULT 10
US-09-960-352-11516
; Sequence 11516, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960.352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11516
; LENGTH: 374
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 49-LIB34-046-Q1-E1-E2
US-09-960-352-11516

Query Match 17.0%; Score 76; DB 10; Length 374;
Best Local Similarity 60.8%; Pred. No. 3.7e-14;
Matches 124; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy 174 GTCAGCCTTTACTGTGAAGCTCAGTGGAAACTTCTCTTCTTCAAGCCCATCATCTT 233
Db 171 GTCTGCCITTTCCGTCAAGCTGAGTGGCCCTTCCAGGACCCTCCAGCCCATTTGTCTT 230
Qy 234 CACAGGGTCCTGTACATGCCAGGAGGATTTAAAGAGGCCATGGGAGTCTTTGCTTG 293
Db 231 CCAGGAAGCTCTGTACAACCATCAGGGCCACTTCGACCCCGCCACTGGTGTGTTCAACTG 290
Qy 294 CAGGTCGCTGGGAATTAATACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 353
Db 291 CAGCGTCCCTGTGTGACCACTTTGGCTTTGACATTTGAGTTGTTTTCAGAGTGTCTCAA 350
Qy 354 TATTTGGCTAATGAGGAAGCAAT 377
Db 351 AGTGTCTTAATGCGGAATGGAAT 374

RESULT 11
US-09-960-352-3684
; Sequence 3684, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960.352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 3684
; LENGTH: 447
; TYPE: DNA
```

```
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 16-LIB34-024-Q1-E1-D7
US-09-960-352-3684

Query Match 16.4%; Score 73.2; DB 10; Length 447;
Best Local Similarity 60.6%; Pred. No. 3.1e-13;
Matches 120; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

Qy 174 GTCAGCCTTTACTGTGAAGCTCAGTGGAAACTTCTCTTCTTCTTCTTCTTCTTCTTCTT 233
Db 160 GTCTGCCITTTCCGTCAAGCTGAGTGGCCCTTCCAGGACCCTCCAGGCCCATTTGTCTT 219
Qy 234 CACAGGGTCCTGTACATGCCAGGAGGATTTAAAGAGGCCATGGGAGTCTTTGCTTG 293
Db 220 CCAGGAAGTCTGTACAACCATCAGGGCCACTTCGACCCCGCCACTGGTGTGTTCAACTG 279
Qy 294 CAGGTCGCTGGGAATTAATACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 353
Db 280 CAGCGTCCCTGTGTGACCACTTTGGCTTTGACATTTGAGTTGTTTTCAGAGTGTCTCAA 339
Qy 354 TATTTGGCTAATGAGGA 371
Db 340 GGTGGTCTAATTCGAA 357

RESULT 12
US-09-960-352-4241
; Sequence 4241, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960.352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 4241
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Bos taurus
; NAME/KEY: unsure
; LOCATION: (377)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 18-LIB34-077-Q1-E1-E5
US-09-960-352-4241

Query Match 13.6%; Score 60.6; DB 10; Length 467;
Best Local Similarity 54.5%; Pred. No. 2.8e-09;
Matches 145; Conservative 0; Mismatches 115; Indels 6; Gaps 1;

Qy 40 CATGTAGAAGTTGCTGGACCTCCAGCACACCACCCAGGCCCCAGAGAAGTGGGCTCTCT 99
Db 204 CTTCCAGGCCCTTCCAGGCCCTCCCGCCCTTCTCTGCCCCCTCTGCGCCCTCCAGGGCCCCA 263
Qy 100 GGTGCACCAGGTTTACCACAATATACAGAGAAATAAGTGAATGACAAAATGCCCCCTGT 159
Db 264 GGCCCTCCAGGATTCGAGGACCAACTGGGATTCGGGAGACATTCAGAGTTGCTGTCT 323
Qy 160 CTTGATATAGAAGTGCAGCCCTTTACTGTGAAGCTCAGTGGAAACTTCTTCTTCTTCTT 219
Db 324 CCCCCAAA-----TCTGCCCTTTCAGTGAAGATGAATGATCCCTCCAGGCCCTCTCN 377
Qy 220 AAGCCCATCATCTTCACAGGGGTCTGTACAATGCCAGAGGATTTTAAAGGAGGCCATG 279
Db 378 CAGCCCATTCCTTCAAGGGAACCTGCATATGACCAGGACCCTTCACTTCACTTACCAGG 437
Qy 280 GGAGTCTTTGCTTGCAGGGTGCCTGG 305
Db 305
```

Db 438 GGAGTGTTCACCTGCACCATCCCTGG 463

RESULT 13

```
US-09-864-761-25080
; Sequence 25080, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25080
; LENGTH: 273
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL121767.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 6.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.5
; OTHER INFORMATION: SWISSPROT HIT: P29539, EVALUE 6.80e-01
; OTHER INFORMATION: EST HUMAN HIT: BE156318.1, EVALUE 3.00e-78
; OTHER INFORMATION: NT HIT: AL163210.2, EVALUE 7.00e-75
US-09-864-761-25080
```

```
Query Match 13.1%; Score 58.4; DB 10; Length 273;
Best Local Similarity 74.0%; Pred. No. 1e-08;
Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 348 GGTGAATATTGGCTAATGAGGAGCAAAATTTTGGCTAATAAGCAAAATTTCTAAGCA 407
Db 115 GGGGAACATTGAACCTTGAGAGAGATGATTTGGGGTATCTCGTGGAGAAATTTCTAAACA 174
QY 408 GCAAAGCATTTCAAGAGGTGACTTTGGGTGCTGTTAAAGGCA 447
Db 175 GCAAAGCATTTCAAGAGGTGACTTTGGCTGCTGTTAAAGGCA 214
```

RESULT 14

```
US-09-864-761-8349
; Sequence 8349, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 8349
; LENGTH: 519
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL121767.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.7
```



THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 16:09:30 ; Search time 29.509 Seconds  
(without alignments)  
4645.518 Million cell updates/sec

Title: US-09-997-610-1\_COPY\_2\_448

Perfect score: 447  
Sequence: 1 atagtggtacactgtctt.....cttggtgtctgttaaaggca 447

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48.4	10.8	1313	2	US-08-463-911-6
2	48.4	10.8	4107	4	US-09-140-804-9
3	42.4	9.5	1107	3	US-09-188-930-217
4	40	8.9	1276	2	US-08-463-911-1
5	37.4	8.4	728	4	US-09-336-536-2
6	37.4	8.4	1338	4	US-09-336-536-1
7	37.4	8.4	1347	4	US-09-140-804-1
8	36.2	8.1	729	4	US-09-140-804-10
9	35.8	8.0	1074	2	US-08-627-151A-15
10	35.8	8.0	1404	6	5171840-8
11	35.8	8.0	1404	6	5480796-8
12	35.8	8.0	1486	4	US-08-795-473B-3
13	35.8	8.0	1486	4	US-09-439-856-3
14	35.8	8.0	2061	6	5171840-1
15	35.8	8.0	2061	6	5480796-1
16	35.8	8.0	3319	4	US-08-795-473B-2
17	35.8	8.0	3319	4	US-09-439-856-2
18	34.4	7.7	14855	2	US-08-687-080-59
19	34.4	7.7	44453	4	US-09-146-053-5
20	34.4	7.7	49136	4	US-09-422-869-1
21	34	7.6	459	4	US-09-415-551-2
22	34	7.6	486	4	US-09-415-551-1
23	33.2	7.4	1881	4	US-09-029-348-20
24	33.2	7.4	168575	4	US-09-426-290-1
25	31.4	7.0	1001	3	US-09-188-930-218
26	31.4	7.0	1015	3	US-09-188-930-30
c 27	31.4	7.0	1395	4	US-09-149-476-88

c 28	31.4	7.0	9388	4	US-08-991-789A-141	Sequence 141, App
c 29	31.4	7.0	9388	4	US-09-062-451-141	Sequence 141, App
c 30	31.4	7.0	9388	4	US-09-598-326-141	Sequence 141, App
c 31	31.4	7.0	35081	2	US-08-752-760A-1	Sequence 1, Appl
c 32	31	6.9	1215	2	US-09-092-770-8	Sequence 8, Appl
c 33	31	6.9	1215	4	US-09-222-851-8	Sequence 8, Appl
c 34	30.8	6.9	3294	4	US-08-923-992A-7	Sequence 7, Appl
c 35	30.6	6.8	2712	1	US-08-346-455B-37	Sequence 37, Appl
c 36	30.6	6.8	2712	3	US-08-977-221-37	Sequence 37, Appl
c 37	30.6	6.8	2712	4	US-09-483-831B-37	Sequence 37, Appl
c 38	30.6	6.8	2712	5	PCT-US95-06613-37	Sequence 35, Appl
c 39	30.6	6.8	2946	1	US-08-346-455B-35	Sequence 35, Appl
c 40	30.6	6.8	2946	3	US-08-977-221-35	Sequence 35, Appl
c 41	30.6	6.8	2946	4	US-09-483-831B-35	Sequence 35, Appl
c 42	30.6	6.8	2946	5	PCT-US95-06613-35	Sequence 35, Appl
c 43	30.6	6.8	3251	1	US-08-346-455B-68	Sequence 68, Appl
c 44	30.6	6.8	3251	3	US-08-977-221-68	Sequence 68, Appl
c 45	30.6	6.8	3251	4	US-09-483-831B-68	Sequence 68, Appl

ALIGNMENTS

RESULT 1  
US-08-463-911-6  
; Sequence 6, Application US/08463911  
; Patent No. 5869330  
; GENERAL INFORMATION:  
; APPLICANT: Scherer, Philipp E.  
; APPLICANT: Lodish, Harvey F.  
; TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED EXCLUSIVELY IN ADIPOCYTES  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/463,911  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia  
; REGISTRATION NUMBER: 32,227  
; REFERENCE/DOCKET NUMBER: WHI95-05  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 861-6240  
; TELEFAX: (617) 861-9540  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1313 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 73..804  
US-08-463-911-6

Query Match 10.8%; Score 48.4; DB 2; Length 1313;  
Best Local Similarity 48.8%; Pred. No. 7.4e-06;  
Matches 164; Conservative 0; Mismatches 166; Indels 6; Gaps 1;  
QY 41 ATGTAGAGTTCTGTGACCTCCAGCACACCCAGGCCCCCAGAGAGTGGGCGCTCTG 100

Db 281 ATCCAGGTCTTATGTTCTTAAGGAGACATCGGTGAACCCGGAGTACCCGGGGCTGAAG 340  
QY 101 GTGACACAGGTTTACACAAATATACAGGAGAAATAGTGAATGACAAAATGCCCTGTC 160  
Db 341 GTCCCCAGGCTTCCGGGAATCCAGGAGGAGAAACCTGGAGAAGTGGCT--- 397  
QY 161 CTGATATAGAAAGGTGAGCTTTACTGTGAAGCTCAGTGGAAACCTCTCTTCTTTCA 220  
Db 398 --ATGATATACCGCTCAGCATTCAGTGTGGATTGGAGACTTACGTTACTATCCCAACA 454  
QY 221 AGCCATCATCTTCAGAGGGTCTGTACAATGCCAGAGGATTTAAAGAGGCGCATGG 280  
Db 455 TGCCCATTCGCTTTACCAAGATCTTACAATCAGCAAAACCACTATGATGGCTCCACTG 514  
QY 281 GAGTCCTTTGCTGAGGGTGCCTGGGAATTACTACTCCAGCTTTGATGTTGAGCTGCATC 340  
Db 515 GTAATTCACCTGCAACATTCCTGGCTGTACTACTTTCCTACCACATCACAGTCTATA 574  
QY 341 ATTGCAAGGTGAATATTTGGCTAATGAGGAAGCAA 376  
Db 575 TGAAGGATGTAAGGTGACGCTCTTCAAGAGGACA 610

RESULT 2  
US-09-140-804-9  
; Sequence 9, Application US/09140804  
; Patent No. 6157930  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.  
; APPLICANT: Humes, Jacqueline M.  
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS  
; FILE REFERENCE: 97-49  
; CURRENT APPLICATION NUMBER: US/09/140,804  
; CURRENT FILING DATE: 1998-08-26  
; EARLIER APPLICATION NUMBER: 60/056,983  
; EARLIER FILING DATE: 1997-08-26  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 9  
; LENGTH: 4517  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-140-804-9  
Query Match 10.8%; Score 48.4; DB 4; Length 4517;  
Best Local Similarity 48.8%; Pred. No. 1.4e-05;  
Matches 164; Conservative 0; Mismatches 166; Indels 6; Gaps 1;

QY 41 ATGTAGAGTTGCTGACCTCCAGCACACCCCGAGGCCCCAGAAAGTGGGGCTCTCTG 100  
Db 235 ATCCAGGTCTTATGTTCTTAAGGAGACATCGGTCAACCGGAGTACCCGGGGCTGAAG 294  
QY 101 GTGACACAGGTTTACACAAATATACAGGAGAAATAGTGAATGACAAAATGCCCTGTC 160  
Db 295 GTCCCCAGGCTTCCGGGAATCCAGGAGGAGAAACCTGGAGAAGTGGCT--- 351  
QY 161 CTGATATAGAAAGGTGAGCTTTACTGTGAAGCTCAGTGGAAACCTCTCTTCTTTCA 220  
Db 352 ---ATGATATACCGCTCAGCATTCAGTGTGGATTGGAGACTTACGTTACTATCCCAACA 408  
QY 221 AGCCATCATCTTCAGAGGGTCCCTGTACAATGCCAGAGGATTTAAAGAGGCGCATGG 280  
Db 409 TGCCCATTCGCTTTACCAAGATCTTCTACAATCAGCAAAACCACTATGATGGCTCCACTG 468  
QY 281 GAGTCCTTTGCTGAGGGTGCCTGGGAATTACTACTCCAGCTTTGATGTTGAGCTGCATC 340  
Db 469 GTAATTCACCTGCAACATTCCTGGGCTGTACTACTTTTCCTACCACATCACAGTCTATA 528  
QY 341 ATTGCAAGGTGAATATTTGGCTAATGAGGAAGCAA 376  
Db 529 TGAAGGATGTAAGGTGACGCTCTTCAAGAGGACA 564

RESULT 3  
US-09-188-930-217  
; Sequence 217, Application US/09188930A  
; Patent No. 6150502  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murison, James Greg  
; TITLE OF INVENTION: Compositions Isolated From Skin Cells  
; TITLE OF INVENTION: and Methods For Their Use  
; FILE REFERENCE: 11000.1011c1  
; CURRENT APPLICATION NUMBER: US/09/188,930A  
; CURRENT FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 348  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 217  
; LENGTH: 1107  
; TYPE: DNA  
; ORGANISM: Rat  
US-09-188-930-217

Query Match 9.5%; Score 42.4; DB 3; Length 1107;  
Best Local Similarity 50.5%; Pred. No. 0.00058;  
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 168 AGAAGGTGAGCTTTACTGTGAAGCTCAGTGGAAACCTTCCCTTCCTTTCAAGCCCAT 227  
Db 614 AGCCAAAGTGGGCTTTTCGTGGCGGTAAACCAAGAGTTACCCACGTGAGCGACTGCCCAT 673  
QY 228 CATCTTCACAGGGTCTCTACAAATGCCAGAGGGATTTAAAGGAGGCGCATGGAGTCTT 287  
Db 674 CAAGTTTGACAAGATTCTGATGAATGAGGAGGCCACTACAATGCATCCAGTGGCAAGTT 733  
QY 288 TGCTTGAGGGTGCCTGGGAATTTACTACTCCAGCTTTGATGTTGAGCTGCATCATTTGCAA 347  
Db 734 CGTCTGAGCGTGCCAGGGATCTATTACTTACCTATGACATTTACGCTGCCCAACAACA 793  
QY 348 GGTGAATATTTGGCTTAATGAGGAA 371  
Db 794 CCTGGCCATCGGCTAGTGACAA 817

RESULT 4  
US-08-463-911-1  
; Sequence 1, Application US/08463911  
; Patent No. 5869330  
; GENERAL INFORMATION:  
; APPLICANT: Scherer, Philipp E.  
; APPLICANT: Lodish, Harvey F.  
; TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED  
; TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/463,911  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia



Db 603 TACGACGCCCTACCGGCAAGTTACCTGCCAGGTGCTTGGGTCTACTACTTGGCGGTC 662  
QY 325 GATGTTGAGCTGATCATTTCAAGGTGAATATTTGGTCAATGAGAA 371  
Db 663 CATGCCACCGCTACCGGGCCAGCCTGCAGTTTCACTGTGTGAAGAA 709

RESULT 8  
US-09-140-804-10  
; Sequence 10, Application US/09140804  
; Patent No. 6197930  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.  
; APPLICANT: Humes, Jacqueline M.  
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS  
; FILE REFERENCE: 97-49  
; CURRENT APPLICATION NUMBER: US/09/140,804  
; CURRENT FILING DATE: 1998-08-26  
; EARLIER APPLICATION NUMBER: 60/056,983  
; EARLIER FILING DATE: 1997-08-26  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 10  
; LENGTH: 729  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Degenerate nucleotide sequence encoding the zsig39  
; OTHER INFORMATION: polypeptide of SEQ ID NO:2.  
US-09-140-804-10

Query Match 8.1%; Score 36.2; DB 4; Length 729;  
Best Local Similarity 27.9%; Pred. No. 0.047;  
Matches 86; Conservative 33; Mismatches 189; Indels 0; Gaps 0;  
QY 64 GCACACCCCGCCGCCAGGAAGTGGGGCTCTCGTGGTGCACACAGGTTTACCACAATAT 123  
Db 205 GGNMGNCNGNNTCCNGCNCNMGGNGAYCCNGCNCNMGGNGARGCNGGNCN 264  
QY 124 ACAGGAAATAAGTGAATAAGCAAAATGCCCTGCTCTGATATAGAAAGTCAGCCTTT 183  
Db 265 GCGNGCNCACNGCNGCNGCNGGNGARTGYWSNGTNCNCNMGNSNGNNTTYSWNGCN 324  
QY 184 ACTGTGAAGCTCAGTGGAAACTTCTCTTCTTCAAGCCCATCATCTTTCACAGGGGTC 243  
Db 325 AARMGNSNGARNSMNGTNCNCNCCNCCNWSNGAYCNCNNTTTCNTTYGAYMNGTN 384  
QY 244 CTGTACATCCCGCAGAGGATTTAAAGGAGGCATGGGAGTCTTTGCTTGCAGGGTGCCCT 303  
Db 385 YTNGTNAAYGARGCNGNCAYTAIGYGCNGTNCNGNNAARTTYACNTGYCARGTNCN 444  
QY 304 GGAATTTACTACCGCTTGTGATGTGAGCTGCATTCATTCGCAAGGTGAATATTTGGCTA 363  
Db 445 GNGTNTAYTYTYGCGTNGTNCAYGNCNAGTNTATMNGCNCNSNTTNCARTTYGAYTN 504  
QY 364 ATGAGGAA 371  
Db 505 GTNAARAA 512

RESULT 9  
US-08-627-151A-15  
; Sequence 15, Application US/08627151A  
; Patent No. 5866341  
; GENERAL INFORMATION:  
; APPLICANT: SPINELLIA, Dominic  
; APPLICANT: BECHERER, Kathleen  
; APPLICANT: BROWN, Steven  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; SCREENING DRUG LIBRARIES  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Gen-Probe Incorporated

; STREET: 10210 Genetic Center Drive  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92121  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/627,151A  
; FILING DATE: 03-APR-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fisher, Carlos A  
; REGISTRATION NUMBER: 36,510  
; REFERENCE/DOCKET NUMBER: CBI016  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-410-8926  
; TELEFAX: 619-410-8928  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1074 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-627-151A-15

Query Match 8.0%; Score 35.8; DB 2; Length 1074;  
Best Local Similarity 54.1%; Pred. No. 0.077;  
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
QY 10 ATACCTGCTCTTAATAACGGCAGTCATGAGCATGTAGAACTTGTGGACCTCCAGCACAC 69  
Db 768 ATATCGGGCTGAACGGTCAAAAGACATTCACAACATGGATGTCGAAGACCTCCAGCATCA 827  
QY 70 CCCAGGCCCCCAGAAGTGGGGCCCTCTGGTGCACAGGTTTACCACAATATACAGGA 129  
Db 828 CTGTGTCTATCCAGCAGCCCTGGAGCGGCTGAGGCGAGTGTGTCGCTGCTGCCAGGA 887  
QY 130 GAAATAGTCAATG 144  
Db 888 GGAGTTCGGGCAAGG 902

RESULT 10  
5171840-8  
; Patent No. 5171840  
; APPLICANT: KISHIMOTO, TADAMITSU  
; TITLE OF INVENTION: RECEPTOR PROTEIN FOR HUMAN B CELL  
; STIMULATORY FACTOR-2  
; NUMBER OF SEQUENCES: 11  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/298,694  
; FILING DATE: 19-JAN-1989  
; SEQ ID NO:8  
; LENGTH: 1404  
5171840-8

Query Match 8.0%; Score 35.8; DB 6; Length 1404;  
Best Local Similarity 54.1%; Pred. No. 0.088;  
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
QY 10 ATACCTGCTCTTAATAACGGCAGTCATGAGCATGTAGAACTTGTGGACCTCCAGCACAC 69  
Db 768 ATATCGGGCTGAACGGTCAAAAGACATTCACAACATGGATGTCGAAGACCTCCAGCATCA 827  
QY 70 CCCAGGCCCCCAGAAGTGGGGCCCTCTGGTGCACAGGTTTACCACAATATACAGGA 129





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 15:08:07 ; Search time 154.704 Seconds  
(without alignments)  
6506.903 Million cell updates/sec

Title: US-09-997-610-1-copy\_2\_448

Perfect score: 447

Sequence: 1 atagtgtcatacctgtctt.....cttggtgtctgttaaaggca 447

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	351	78.5	1338	24	Human genset metab
2	58.6	13.1	404	21	Human secreted pro
3	58.4	13.1	273	22	Human foetal liver
4	58.4	13.1	273	22	Probe #18226 for g
5	58.4	13.1	273	22	Human brain expres
6	58.4	13.1	273	22	Human bone marrow
7	58.4	13.1	273	22	Probe #16759 for g
8	58.4	13.1	273	22	Probe #24296 used
9	58.4	13.1	519	22	Human foetal liver

10	58.4	13.1	519	22	ABA29883	Probe #8349 for ge
11	58.4	13.1	519	22	AAK10911	Human brain expres
12	58.4	13.1	519	22	AAK36769	Human bone marrow
13	58.4	13.1	519	22	AAI17621	Probe #7554 for ge
14	58.4	13.1	519	22	AAI42541	Probe #11227 used
15	57	12.8	662	22	AAI5063	Human breast cance
16	57	12.8	686	22	AAI23916	Human breast cance
17	57	12.8	930	23	AAI81134	DNA encoding novel
18	57	12.8	999	23	AAI66807	DNA encoding novel
19	57	12.8	1077	23	AAI82659	DNA encoding novel
20	57	12.8	1946	22	ABA44227	Human breast cell
21	57	12.8	1946	22	ABA54677	Human foetal liver
22	57	12.8	1946	22	ABA24459	Probe #2925 for ge
23	57	12.8	1946	22	AAK02964	Human brain expres
24	57	12.8	1946	22	AAK28410	Human bone marrow
25	57	12.8	1946	22	AAI12975	Probe #2908 for ge
26	57	12.8	1946	22	AAI34334	Probe #3020 used t
27	57	12.8	1946	22	AAI02893	Probe #2884 used t
28	57	12.8	1946	22	ABS02918	Human genome-deriv
c 29	56.4	12.6	32127	24	AAI99255	Human excretory re
c 30	56.4	12.6	32127	22	AAI63605	Human kidney relat
c 31	56.4	12.6	72215	22	AAK86832	Human immune/haema
32	56	12.5	505	22	ABA60370	Human foetal liver
33	56	12.5	505	22	ABA28609	Probe #7075 for ge
34	56	12.5	505	22	AAK08648	Human brain expres
35	56	12.5	505	22	AAK34535	Human bone marrow
36	56	12.5	505	22	AAI40252	Probe #8938 used t
37	56	12.5	505	24	ABS09176	Human genome-deriv
38	56	12.5	1054	23	ABS64468	DNA encoding novel
c 39	55.4	12.4	342	21	AAK26950	Human secreted pro
40	55.4	12.4	520	22	ABA62906	Human foetal liver
41	55.4	12.4	520	22	ABA30177	Probe #8643 for ge
42	55.4	12.4	520	22	AAK11311	Human brain expres
43	55.4	12.4	520	22	AAK37101	Human bone marrow
44	55.4	12.4	520	22	AAI17942	Probe #7875 for ge
45	55.4	12.4	520	22	AAI42928	Probe #11614 used

ALIGNMENTS

RESULT 1	
AAI44066	
ID	AAI44066 standard; cDNA; 1338 BP.
XX	
AC	AAI44066;
XX	
XX	
DT	27-SEP-2002 (first entry)
DE	Human genset metabolic gene (GMG-9) cDNA sequence.
XX	
KW	Human; gene; ss; gene therapy; genset metabolic gene; GMG-7A; GMG-7B;
KW	GMG-8; GMG-9; GMG-10; GMG-11; metabolic-related disorder; obesity;
KW	impaired glucose tolerance; insulin resistance; Syndrome X;
KW	Type II diabetes; hyperlipidaemia; atherosclerosis; hypertension;
KW	heart disease; cardiac insufficiency; coronary insufficiency;
KW	high blood pressure; insulin sensitizer;
KW	non-insulin dependent diabetes mellitus.
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
CDS	1..1338
FT	/*tag= a
FT	/partial
FT	/product= "Human GMG-9 protein"
XX	/note= "No stop codon is given"
XX	
XX	WO200255694-A2.
PD	18-JUL-2002.
XX	
PF	15-JAN-2002; 2002WO-IB01215.



```
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX Claim 4; SEQ ID NO 23381; 639pp + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 273 BP; 80 A; 36 C; 81 G; 76 T; 0 other;
Query Match 13.1%; Score 58.4; DB 22; Length 273;
Best Local Similarity 74.0%; Pred. No. 1.2e-08;
Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY 348 GGTGAATATTTGGCTAATGAGGAGCAAAATTTGGCTAATGAGGAGCAAAATTTCTAAGCA 407
DB 115 GGGGAACATTTGACCTTGAGAGAGATGATTTGGGGTATCTGGTGGAGAGAAATTTCTAAGCA 174
QY 408 GCAAGAGCATTCAGAGGCTGACTTGGCTGCTGTTAAAGGCA 447
DB 175 GCAAGAGCATTCAGAGGCTGACTTGGCTGCTGTTAAAGGCA 214
RESULT 4
ABA39760
ID ABA39760 standard; DNA; 273 BP.
XX AC ABA39760;
XX DT 23-JAN-2002 (first entry)
XX DE Probe #18226 for gene expression analysis in human heart cell sample.
XX KW Human; gene expression; heart; microarray; vascular system; probe;
XX KW cardiovascular disease; hypertension; cardiac arrhythmia;
XX KW congenital heart disease; ss.
XX OS Homo sapiens.
XX PN WO200157274-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00666.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
```

```
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX Claim 4; SEQ ID NO 18226; 530pp; English.
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 273 BP; 80 A; 36 C; 81 G; 76 T; 0 other;
Query Match 13.1%; Score 58.4; DB 22; Length 273;
Best Local Similarity 74.0%; Pred. No. 1.2e-08;
Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY 348 GGTGAATATTTGGCTAATGAGGAGCAAAATTTGGCTAATGAGGAGCAAAATTTCTAAGCA 407
DB 115 GGGGAACATTTGACCTTGAGAGAGATGATTTGGGGTATCTGGTGGAGAGAAATTTCTAAGCA 174
QY 408 GCAAGAGCATTCAGAGGCTGACTTGGCTGCTGTTAAAGGCA 447
DB 175 GCAAGAGCATTCAGAGGCTGACTTGGCTGCTGTTAAAGGCA 214
RESULT 5
AAK23591
ID AAK23591 standard; DNA; 273 BP.
XX AC AAK23591;
XX DT 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe SEQ ID NO: 23582.
XX KW Human; brain expressed exon; gene expression analysis; probe;
XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX KW epilepsy; cancer; ss.
XX OS Homo sapiens.
XX PN WO200157275-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00667.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
```

XX Single exon nucleic acid probes for analyzing gene expression in human  
PT brains -  
XX  
XX Example 4; SEQ ID NO: 23582; 650pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention.  
XX  
XX Sequence 273 BP; 80 A; 36 C; 81 G; 76 T; 0 other;  
SQ  
Query Match 13.1%; Score 58.4; DB 22; Length 273;  
Best Local Similarity 74.0%; Pred. No. 1.2e-08;  
Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
QY 348 GGTGAATATTGGCTAATGAGGAGCAAAATTTGGCTAATAAGGAAGAAATTTCTTAAGCA 407  
DB 115 GGGACATTTGAACTTGAGAGAGATGATTTGGGGTATCTGGTGGAGAAATTTCTTAACA 174  
QY 408 GCAAGCATTTCAAGAGGTGACTTGGTGGTGTGTTAAAGGCA 447  
DB 175 GCAAGCATTTCAAGAGGTGACTTGGTGGTGTGTTAAAGGCA 214  
RESULT 6  
AAK49719  
ID AAK49719 standard; DNA; 273 BP.  
XX  
AC AAK49719;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed single exon probe SEQ ID NO: 24276.  
XX  
DE Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00668.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488900/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human bone marrow -  
XX  
XX Example 4; SEQ ID NO: 24276; 658pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers

CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
CC the probes of the invention.  
XX  
XX Sequence 273 BP; 80 A; 36 C; 81 G; 76 T; 0 other;  
SQ  
Query Match 13.1%; Score 58.4; DB 22; Length 273;  
Best Local Similarity 74.0%; Pred. No. 1.2e-08;  
Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
QY 348 GGTGAATATTGGCTAATGAGGAGCAAAATTTGGCTAATAAGGAAGAAATTTCTTAAGCA 407  
DB 115 GGGACATTTGAACTTGAGAGAGATGATTTGGGGTATCTGGTGGAGAAATTTCTTAACA 174  
QY 408 GCAAGCATTTCAAGAGGTGACTTGGTGGTGTGTTAAAGGCA 447  
DB 175 GCAAGCATTTCAAGAGGTGACTTGGTGGTGTGTTAAAGGCA 214  
RESULT 7  
AAI26826  
ID AAI26826 standard; DNA; 273 BP.  
XX  
AC AAI26826;  
XX  
DT 12-OCT-2001 (first entry)  
XX  
DE Probe #16759 for gene expression analysis in human cervical cell sample.  
XX  
DE Probe; human; microarray; gene expression; cervical epithelial cell;  
KW cervical cancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157278-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00670.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488901/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human cervical epithelial cells -  
XX  
XX Claim 25; SEQ ID No 16759; 487pp; English.  
XX  
CC The present invention relates to human single exon nucleic acid probes  
CC (SENP). The present sequence is one such probe. The SENPs are derived  
CC from human HeLa cells. The SENPs can be used to produce a single exon  
CC microarray, which can be used for measuring human gene expression in a  
CC sample derived from human cervical epithelial cells. By measuring gene  
CC expression, the probes are therefore useful in grading and/or staging  
CC of diseases of the cervix, notably cervical cancer.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 273 BP; 80 A; 36 C; 81 G; 76 T; 0 other;  
SQ  
Query Match 13.1%; Score 58.4; DB 22; Length 273;  
Best Local Similarity 74.0%; Pred. No. 1.2e-08;



```
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease; ss.
OS Homo sapiens.
XX
XX WO200157274-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00666.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts .
XX
XX Claim 1; SEQ ID No 8349; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 519 BP; 170 A; 68 C; 148 G; 133 T; 0 other;

Query Match          13.1%; Score 58.4; DB 22; Length 519;
Best Local Similarity 74.0%; Pred. No. 1.7e-08;
Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 348 GGTGAATATTGGCTAATGAGGAGCAAAATTTTGGCTAATAAGGAAATTTCTAAGCA 407
DB 254 GGGGACATTTGAACCTTGAGAGAGATGATTGGGGTATCTGGTGAAGAAATTTCTAACA 313

QY 408 GCAAAGCATTTCAAGAGGTGACTTGGTGGTCTGTTAAAGGCA 447
DB 314 GCAAAGCATTTCAAGAGGTGACTTGGTGGTCTGTTAAAGGCA 353

RESULT 11
AAK10911
ID AAK10911 standard; DNA; 519 BP.
XX
XX AAK10911;
XX
XX 05-NOV-2001 (first entry)
XX
XX Human brain expressed single exon probe SEQ ID NO: 10902.
XX
XX Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer; ss.
XX
```

```
OS Homo sapiens.
XX
XX WO200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains .
XX
XX Example 4; SEQ ID NO: 10902; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
XX
XX Sequence 519 BP; 170 A; 68 C; 148 G; 133 T; 0 other;

Query Match          13.1%; Score 58.4; DB 22; Length 519;
Best Local Similarity 74.0%; Pred. No. 1.7e-08;
Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 348 GGTGAATATTGGCTAATGAGGAGCAAAATTTTGGCTAATAAGGAAATTTCTAAGCA 407
DB 254 GGGGACATTTGAACCTTGAGAGAGATGATTGGGGTATCTGGTGAAGAAATTTCTAACA 313

QY 408 GCAAAGCATTTCAAGAGGTGACTTGGTGGTCTGTTAAAGGCA 447
DB 314 GCAAAGCATTTCAAGAGGTGACTTGGTGGTCTGTTAAAGGCA 353

RESULT 12
AAK36769
ID AAK36769 standard; DNA; 519 BP.
XX
XX AAK36769;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human bone marrow expressed single exon probe SEQ ID NO: 11326.
XX
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
XX Homo sapiens.
XX
XX WO200157276-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00668.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX
```

```
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 11326; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukemia and myeloma. The present sequence is one of
XX the probes of the invention.
XX
XX Sequence 519 BP; 170 A; 68 C; 148 G; 133 T; 0 other;
XX
XX Query Match 13.1%; Score 58.4; DB 22; Length 519;
XX Best Local Similarity 74.0%; Pred. No. 1.7e-08;
XX Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
XX
QY 348 GGTGAATATTTGGCTAATGAGAGCAAAATTTGGCTAATGAGAGCAAAATTTCTTAAGCA 407
DB ||||| || || || || || || || || || || || || || || || || || || || ||
254 GGGGAACATTGAACCTTGAGAGAGATGATTTGGGTATCTGTTGAAGAAATTTCTTAACA 313
QY 408 GCAAGCAATTCAGAGGTGACTTGGTCTGTTAAAGCA 447
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
314 GCAAGCAATTCAGAGGTGACTTGGTCTGTTAAAGCA 353

RESULT 13
AA117621
ID AA117621 standard; DNA; 519 BP.
XX
XX AC AA117621;
XX
XX 12-OCT-2001 (first entry)
XX
XX Probe #7554 for gene expression analysis in human cervical cell sample.
XX
XX Probe; human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157278-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488901/53.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs).
```

```
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 25; SEQ ID No 7554; 487pp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENPs). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging
XX of diseases of the cervix, notably cervical cancer.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 519 BP; 170 A; 68 C; 148 G; 133 T; 0 other;
XX
XX Query Match 13.1%; Score 58.4; DB 22; Length 519;
XX Best Local Similarity 74.0%; Pred. No. 1.7e-08;
XX Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
XX
QY 348 GGTGAATATTTGGCTAATGAGAGCAAAATTTGGCTAATGAGAGCAAAATTTCTTAAGCA 407
DB ||||| || || || || || || || || || || || || || || || || || || || ||
254 GGGGAACATTGAACCTTGAGAGAGATGATTTGGGTATCTGTTGAAGAAATTTCTTAACA 313
QY 408 GCAAGCAATTCAGAGGTGACTTGGTCTGTTAAAGCA 447
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
314 GCAAGCAATTCAGAGGTGACTTGGTCTGTTAAAGCA 353

RESULT 14
AA142541
ID AA142541 standard; DNA; 519 BP.
XX
XX AC AA142541;
XX
XX 17-OCT-2001 (first entry)
XX
XX Probe #11227 used to measure gene expression in human placenta sample.
XX
XX Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488997/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID No 11227; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs).
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:16:26 ; Search time 1425.16 Seconds  
(without alignments)  
9128.055 Million cell updates/sec

Title: US-09-997-610-1\_COPY\_2\_448

Perfect score: 447

Sequence: 1 atagtgatcatactgtctt.....cttggtgctgttaaggca 447

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg\_hum.\*

31: em.htg\_inv.\*

32: em.htg\_other.\*

33: em.htg\_mus.\*

34: em.htg\_pln.\*

35: em.htg\_rod.\*

36: em.htg\_mam.\*

37: em.htg\_vrt.\*

38: em\_sy.\*

39: em.htgo\_hum.\*

40: em.htgo\_mus.\*

41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	351	78.5	145880	9	HS302D9	282198 Human DNA s
2	151.2	33.8	1265	10	TMSHP20A	DI2974 Tamias asia
3	139.2	31.1	5121	10	AB067779	AB067779 Tamias si
4	84.6	18.9	2005	10	TMSHP25	DI2975 Tamias asia
5	68.6	15.3	559	10	AB067813S3	AB069679 Callosciu
6	68.2	15.3	145880	9	HS302D9	282198 Human DNA s
7	63.4	14.2	90497	9	HS73284	AL008722 Human DNA
8	63.4	14.2	135880	2	AC104256	AC104256 Homo sapi
9	63.2	14.1	51000	9	AP005265	AP005265 Homo sapi
10	63.2	14.1	144514	2	AC027433	AC027433 Homo sapi
11	63.2	14.1	149038	2	AC025112	AC025112 Homo sapi
12	62.6	14.0	162063	9	AC015553	AC015553 Homo sapi
13	62.6	14.0	175134	2	AL161623	AL161623 Homo sapi
14	62.6	14.0	176932	9	AL157884	AL157884 Human DNA
15	62.4	14.0	1385	10	TMSHP27	DI2976 Tamias asia
16	62	13.9	203234	9	AL441943	AL441943 Human DNA
17	61.8	13.8	69937	2	AC069566	AC069566 Homo sapi
18	61.8	13.8	122231	2	AC026099	AC026099 Homo sapi
19	61.8	13.8	143872	9	AC091987	AC091987 Homo sapi
20	61.8	13.8	172792	9	AC062031	AC062031 Homo sapi
21	61.8	13.8	179848	9	AC092967	AC092967 Homo sapi
22	61.8	13.8	180038	2	AC026870	AC026870 Homo sapi
23	61.2	13.7	177386	2	AC018971	AC018971 Homo sapi
24	60.2	13.5	108511	9	AC011418	AC011418 Homo sapi
25	60.2	13.5	137591	2	AL672086	AL672086 Homo sapi
26	60.2	13.5	158587	9	AC074325	AC074325 Homo sapi
27	60.2	13.5	158811	2	AC013503	AC013503 Homo sapi
28	60.2	13.5	160151	9	AC098595	AC098595 Homo sapi
29	60.2	13.5	161049	2	AL591842	AL591842 Homo sapi
30	60.2	13.5	173767	9	AC002449	AC002449 Human PAC
31	60.2	13.5	175765	9	AC027128	AC027128 Homo sapi
32	60.2	13.5	179484	9	AC107626	AC107626 Homo sapi
33	60.2	13.5	187201	2	AC068063	AC068063 Homo sapi
34	60.2	13.5	187640	9	AC073840	AC073840 Homo sapi
35	60.2	13.5	188622	9	AL359258	AL359258 Human DNA
36	60.2	13.5	192973	9	AC023471	AC023471 Homo sapi
37	60.2	13.5	207600	9	AC024910	AC024910 Homo sapi
38	60.2	13.5	222542	9	AC022379	AC022379 Homo sapi
39	59.8	13.4	170200	9	AC008012	AC008012 Homo sapi
40	59.6	13.3	144057	2	AC074266	AC074266 Homo sapi
41	59.6	13.3	172349	9	AC019185	AC019185 Homo sapi
42	59.6	13.3	195594	2	AC096992	AC096992 Homo sapi
43	59.6	13.3	197078	2	AC117382	AC117382 Homo sapi
44	59.6	13.3	233877	9	AC093798	AC093798 Homo sapi
45	59.4	13.3	98070	9	AC025463	AC025463 Homo sapi

ALIGNMENTS

RESULT 1

HS302D9

LOCUS

DEFINITION

Human DNA sequence from clone RPL-302D9 on chromosome 22 Contains

GSSs, complete sequence.

ACCESSION 282198

VERSION 282198.2

KEYWORDS HTG.

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 145880)

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE Bridgeman,A. Direct Submission

linear PRI 12-DEC-1999

## JOURNAL

Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk

On Dec 13, 1999 this sequence version replaced gi:3164067. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>

RP1-302D9 is from the library RPci-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR: PCYPAC2>

This sequence is the entire insert of clone RP1-302D9 The true left end of clone CTA-282F2 is at 69682 in this sequence. The true right end of clone CTA-415G2 is at 55167 in this sequence.

## FEATURES

Source	Location/Qualifiers
1..145980	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="22"
	/clone="RP1-302D9"
	/clone_lib="RPci-1"
repeat_region	188..245
	/note="MER3 repeat: matches 144..209 of consensus"
repeat_region	246..571
	/note="AluX repeat: matches 1..312 of consensus"
repeat_region	572..759
	/note="MER3 repeat: matches 1..144 of consensus"
repeat_region	783..933
	/note="MER5A repeat: matches 26..187 of consensus"
repeat_region	1033..1336
	/note="AluSp repeat: matches 1..299 of consensus"
repeat_region	1450..1583
	/note="MIR repeat: matches 24..160 of consensus"
repeat_region	1687..1752
	/note="L2 repeat: matches 2593..2661 of consensus"
repeat_region	2350..2660
	/note="AluSc repeat: matches 3..309 of consensus"
repeat_region	2684..2981
	/note="AluSq repeat: matches 2..300 of consensus"
repeat_region	3323..3343
	/note="MLT1E repeat: matches 116..136 of consensus"
repeat_region	3344..3652
	/note="AluY repeat: matches 1..309 of consensus"
repeat_region	3653..3928
	/note="MLT1E repeat: matches 136..359 of consensus"
repeat_region	3929..4278
	/note="RHE1B repeat: matches 3..364 of consensus"
repeat_region	4279..4485
	/note="MLT1E repeat: matches 359..568 of consensus"
repeat_region	5073..5176
	/note="52 copies 2 mer ct 78 conserved"
repeat_region	5181..5491
	/note="AluJb repeat: matches 1..311 of consensus"
repeat_region	6369..6485
	/note="L2 repeat: matches 2579..2705 of consensus"
repeat_region	6647..6685
	/note="MADE1 repeat: matches 1..23 of consensus"
repeat_region	6686..6987
	/note="AluX repeat: matches 1..302 of consensus"
repeat_region	6988..7036
	/note="MADE1 repeat: matches 23..77 of consensus"
repeat_region	7482..7754
	/note="AluJb repeat: matches 9..290 of consensus"
repeat_region	7775..8060
	/note="AluJo repeat: matches 1..295 of consensus"
repeat_region	8414..8551
	/note="L2 repeat: matches 2553..2706 of consensus"
repeat_region	8914..9030
	/note="MIR repeat: matches 147..262 of consensus"
repeat_region	9110..9280
	/note="MIR repeat: matches 91..262 of consensus"
repeat_region	9283..9412
	/note="MIR repeat: matches 15..144 of consensus"
repeat_region	9521..9679
	/note="FAM repeat: matches 3..161 of consensus"
repeat_region	9820..10225
	/note="MS1B repeat: matches 2..425 of consensus"
misc_feature	complement(10179..10678)
	/note="match: GSS: Em:B56592"
misc_feature	complement(10204..10728)
	/note="match: GSS: Em:AQ701486"
misc_feature	complement(10249..10706)
	/note="match: GSS: Em:AQ225495"
repeat_region	10312..10383
	/note="MIR repeat: matches 79..150 of consensus"
misc_feature	10718..11310
	/note="match: GSS: Em:B14024"
misc_feature	10784..11201
	/note="match: GSS: Em:B43656"
repeat_region	11838..11946
	/note="MIR repeat: matches 20..137 of consensus"
repeat_region	12174..12445
	/note="L2 repeat: matches 1988..2275 of consensus"
repeat_region	12444..12642
	/note="MIR repeat: matches 63..241 of consensus"
misc_feature	13017..13369
	/note="match: STS: Em:G49301"
repeat_region	13331..13397
	/note="MIR repeat: matches 174..244 of consensus"
repeat_region	13398..13698
	/note="AluSp repeat: matches 1..302 of consensus"
repeat_region	13699..13810
	/note="MIR repeat: matches 76..174 of consensus"
repeat_region	13806..13919
	/note="MIR repeat: matches 77..189 of consensus"
repeat_region	13945..14060
	/note="MIR repeat: matches 24..142 of consensus"
repeat_region	14061..14367
	/note="AluY repeat: matches 1..301 of consensus"
repeat_region	14368..14452
	/note="MIR repeat: matches 141..225 of consensus"
repeat_region	14589..14679
	/note="MIR repeat: matches 173..262 of consensus"
misc_feature	14597..15201
	/note="match: GSS: Em:AQ553482"
misc_feature	14616..15060
	/note="match: GSS: Em:AQ370601"
repeat_region	14868..15040
	/note="MIR repeat: matches 49..233 of consensus"
repeat_region	15071..15188
	/note="L2 repeat: matches 2112..2239 of consensus"
repeat_region	15304..15399
	/note="MLT1B repeat: matches 1..99 of consensus"
repeat_region	15490..15662
	/note="AluSg1 repeat: matches 2..114 of consensus"
repeat_region	15669..15727
	/note="MLT1B repeat: matches 119..178 of consensus"
repeat_region	15728..16027

repeat_region	/note="AluSc repeat: matches 1. .299 of consensus" 16028. .16245
repeat_region	/note="MT1B repeat: matches 178. .390 of consensus" 16546. .16854
repeat_region	/note="AluY repeat: matches 1. .300 of consensus" 18296. .18373
repeat_region	/note="M5TA repeat: matches 2. .29 of consensus" 18324. .18392
repeat_region	/note="MER66-internal repeat: matches 4919. .4993 of consensus" 18393. .18712
repeat_region	/note="AluJb repeat: matches 1. .311 of consensus" 18713. .19133
misc_feature	/note="MER66-internal repeat: matches 4548. .4919 of consensus" complement(18872. .19230)
misc_feature	/note="match: GSS: Em:AQ005063" 19251. .19719
repeat_region	/note="match: GSS: Em:B14179" 19337. .20290
repeat_region	/note="HERVFB21 repeat: matches 4657. .5784 of consensus" 20317. .20382
repeat_region	/note="33 copies 2 mer ta 68 conserved" 20513. .20666
repeat_region	/note="77 copies 2 mer tt 70 conserved" 20682. .21008
repeat_region	/note="AluSg1 repeat: matches 1. .306 of consensus" 21239. .21553
repeat_region	/note="HUES-p3 repeat: matches 4410. .4713 of consensus" 21882. .22254
repeat_region	/note="THE1B repeat: matches 1. .364 of consensus" 22302. .22537
repeat_region	/note="MER66-internal repeat: matches 2186. .2417 of consensus" 22538. .22850
repeat_region	/note="AluSp repeat: matches 1. .313 of consensus" 22851. .23801
repeat_region	/note="MER66-internal repeat: matches 1210. .2186 of consensus" 23905. .23989
repeat_region	/note="MER66-internal repeat: matches 3017. .3102 of consensus"

Query Match	78.5%	Score 351;	DB 9;	Length 145880;	
Best Local Similarity	98.6%;	pred. No. 3.8e-93;			
Matches 354;	Conservative	0;	Mismatches 5;	Indels 0;	Gaps 0;
QY	89	TGGGGCCTCCTCGTGCACCAAGTTTACCACAAATATACAGGAGAAATAGTGAATGACAA	148		
Db	37254	TGGAATGCTCTGGTGGCCCCAGGTTTACCACAAATATACAGGAGAAATAGTGAATGACAA	37313		
QY	149	AATGCCCTCTGCTCGATATAGAAAGGTCAGGCTTTACTGTGAAGCTCAGTGGAAAACTTC	208		
Db	37314	AATGCCCTCTGCTGATATAGAAAGGTCAGGCTTTACTGTGAAGCTCAGTGGAAAACTTC	37373		
QY	209	CTCTTCCTTTCAAGCCCATCATCTTCACAGGGTCCCTGTACAAATGCCAGAGGATTTAA	268		
Db	37374	CTCTTCCTTTCAAGCCCATCATCTTCACAGGGTCCCTGTACAAATGCCAGAGGATTTAA	37433		
QY	269	AGGAGGCCATGGGAGTCTTTGCTTGCAGGTCCTGGGAATTTACTACTCCAGCTTTGATG	328		
Db	37434	AGGAGGCCATGGGAGTCTTTGCTTGCAGGTCCTGGGAATTTACTACTCCAGCTTTGATG	37493		
QY	329	TTGAGCTTGCATCATTTGCAAGGTGAATATTTTGGCTTAATCAGGAAGCAAAATTTGGCTAATA	388		
Db	37494	TTGAGCTTGCATCATTTGCAAGGTGAATATTTTGGCTTAATCAGGAAGCAAAATTTGGCTAATA	37553		
QY	389	AGGAAGAAATTTCAAGCAGCAAGACATTCOAAGGTGACTTTGGGTGCTGTTAAAGGCA	447		
Db	37554	AGGAAGAAATTTCAAGCAGCAAGACATTCOAAGGTGACTTTGGGTGCTGTTAAAGGCA	37612		

RESULT 2  
TMSHP20A

[illegible]

[illegible]

Qy	222	CCCCATCATCTTCACAGGGGTCCTGTACAAATGCCACAGAGGATTTAAAGGAGGCCATGGG	281
Db	428	CCCCATTGCTTCAAGGAGGCTGTGTACAATCAGGAGGGCCACATTCACATGGCCAGGG	487
Qy	282	AGTCTTTGCTTGACAGGTGCCTCGGAATTTACTACTCAGCTTTTCATGCTTGAGCTGCATCA	341
Db	488	AGAGTTCAGCTGTCTCCCTCGTGTGTACAAATTTGGCTTTGCATTCGACTGTTTCA	547
Qy	342	TTGCAAGGTGAATATTTGGCTTAATCAGGAAGCAAAATTTGGCTTAATAGGAAGAA	396
Db	548	GAGTTCTGTGAAGATCAGGCTCATGAGGATGGTATCCAGGTCAGAGAGAAGAA	602
RESULT 5			
LOCUS	AB067813S3	559 bp DNA linear	ROD 18-JAN-2002
DEFINITION	Callosclurus caniceps HP-25 pseudogene, exon 3, complete cds.		
ACCESSION	AB069679		
VERSION	AB069679.1	GI:18149913	
KEYWORDS	3 of 3		
SEGMENT			
SOURCE	Callosclurus caniceps DNA, clone:lambda TS25G1.		
ORGANISM	Callosclurus caniceps		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Sciurinae;		
	Callosclurus.		
REFERENCE	1		
AUTHORS	Kojima, M., Shiba, T., Kondo, N. and Takamatsu, N.		
TITLE	The tree squirrel HP-25 gene is a pseudogene		
JOURNAL	Eur. J. Biochem. 268 (22), 5997-6002 (2001)		
MEDLINE	21579740		
REFERENCE	2 (bases 1 to 559)		
AUTHORS	Takamatsu, N.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-AUG-2001) Nobuhiko Takamatsu, Kitasato University,		
	Department of Biosciences, School of Science; 1-15-1 Kitasato,		
	Sagamihara, Kanagawa 228-8555, Japan		
	(E-mail: takamatu@jet.sci.kitasato-u.ac.jp, Tel: 81427789408,		
	Fax: 81427789408)		
FEATURES	Location/Qualifiers		
source	1..559		
	/organism="Callosclurus caniceps"		
	/db_xref="taxon:64664"		
gene	/clone="lambda TS25G1"		
	join(AB067813, 1:1503..1658, AB069678, 1:26..185, 39..559)		
	/gene="HP-25"		
CDS	join(AB067813, 1:1503..1658, AB069678, 1:26..185, 39..496)		
	/gene="HP-25"		
	/pseudo		
	/codon_start=1		
	/product="HP-25"		
exon	39..>559		
	/gene="HP-25"		
	/number=3		
	/pseudo		
BASE COUNT	150 a 128 c 141 g 140 t		
ORIGIN			
Query Match	15.3%; Score 68.6; DB 10; Length 559;		
Best Local Similarity	55.5%; Pred. No. 2.6e-09;		
Matches	156; Conservative 0; Mismatches 119; Indels 6; Gaps 1;		
Qy	91	GGGCCTCCTGGTCACAGGTTTACCACATATACAGGAGANAATAAGTGAATCACAATA	150
Db	38	GGACATCCAGGTATACCAGGAAGGCCAGGATGACGAGGCCCTCTCTGGAGACATGGAGCG	97
Qy	151	TGCCCTGTCTGTATAGAAAGGTCCAGCCTTTACTGTGAAGCTCAGTGGGAAAACCTTCCT	210
Db	98	TGCCCATCACG-----ACCAAAATCGCCTTTGCCGTGAAGCTGAGTGAGCGTCCCCCA	151
Qy	211	CTTCCTTTCAAGGCCATCATCTTTCACAGGGGTCTCTGTACAAATGCCACGAGGGATTTAAAG	270

Db	152	GGGTCCTTCAGCCCATCATATTCACGAGTCTCTGTACAAACCGGAGGACCACCTATAAC	211
Qy	271	GAGGCCATGGGAGTCTTTTGGCTTGCAGGTCCTCGGGAATTACTACTCCAGCTTTGATGTT	330
Db	212	GTGACCTACTGGAGAGTTTCAGCTATACCAACCCCGTGTGTACAATTTTGGCTTTGACATT	271
Qy	331	GAGCTGCATCATTCGACAGGTGCAATATTTTGGCTAATGAGGAA	371
Db	272	GGACTGTTTCAGAGTCTCTGTGAAGATAAGTCTCATGAGGAA	312
RESULT 6			
HS302D9/c			
LOCUS	Human DNA sequence from clone RP1-302D9 on chromosome 22 Contains		
DEFINITION	GSs, complete sequence.		
ACCESSION	282198		
VERSION	282198.2 GI:6572207		
KEYWORDS	HTG.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 145880)		
JOURNAL	Direct Submission		
COMMENT	Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Dec 13, 1999 this sequence version replaced gi:3164067. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: En.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C-elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr22 RP1-302D9 is from the library RPCL-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2 This sequence is the entire insert of clone RP1-302D9 The true left end of clone CTA-282F2 is at 69682 in this sequence. The true right end of clone CTA-415G2 is at 55167 in this sequence.		
FEATURES	Location/Qualifiers		
source	1. .145880 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="22" /clone="RP1-302D9" /clone_lib="RPCL-1" 188. .245 /note="MER3 repeat: matches 144. .209 of consensus" 246. .571 /note="AluX repeat: matches 1. .312 of consensus" 572. .759 /note="MER3 repeat: matches 1. .144 of consensus" 783. .933 /note="MER5A repeat: matches 26. .187 of consensus"		
repeat_region			
repeat_region			
repeat_region			
repeat_region			

```
repeat_region 1033..1336 /note="AluSp repeat: matches 1. .299 of consensus"  
repeat_region 1450..1583 /note="MIR repeat: matches 24. .160 of consensus"  
repeat_region 1687..1752 /note="MIR repeat: matches 24. .189 of consensus"  
repeat_region 2350..2660 /note="L2 repeat: matches 2593. .2661 of consensus"  
repeat_region 2684..2981 /note="AluSc repeat: matches 3. .309 of consensus"  
repeat_region 3323..3343 /note="AluSq repeat: matches 2. .300 of consensus"  
repeat_region 3344..3652 /note="MLTIE repeat: matches 116. .136 of consensus"  
repeat_region 3653..3928 /note="AluY repeat: matches 1. .309 of consensus"  
repeat_region 3929..4278 /note="MLTIE repeat: matches 136. .359 of consensus"  
repeat_region 4279..4485 /note="HEIB repeat: matches 3. .364 of consensus"  
repeat_region 5073..5176 /note="MLTIE repeat: matches 359. .568 of consensus"  
repeat_region 5181..5491 /note="52 copies 2 mer ct 78 conserved"  
repeat_region 6369..6485 /note="AluJb repeat: matches 1. .311 of consensus"  
repeat_region 6647..6685 /note="L2 repeat: matches 2579. .2705 of consensus"  
repeat_region 6686..6987 /note="WADE1 repeat: matches 1. .23 of consensus"  
repeat_region 6988..7036 /note="AluX repeat: matches 1. .302 of consensus"  
repeat_region 7482..7754 /note="WADE1 repeat: matches 23. .77 of consensus"  
repeat_region 7775..8060 /note="AluJb repeat: matches 9. .290 of consensus"  
repeat_region 8414..8551 /note="AluJo repeat: matches 1. .295 of consensus"  
repeat_region 8914..9030 /note="L2 repeat: matches 2553. .2706 of consensus"  
repeat_region 9110..9280 /note="MIR repeat: matches 147. .262 of consensus"  
repeat_region 9283..9412 /note="MIR repeat: matches 91. .262 of consensus"  
repeat_region 9521..9679 /note="MIR repeat: matches 15. .144 of consensus"  
repeat_region 9820..10225 /note="FAM repeat: matches 3. .161 of consensus"  
misc_feature /note="MSTB repeat: matches 2. .425 of consensus"  
misc_feature /note="match: GSS: Em:B56592"  
misc_feature /note="match: GSS: Em:A0701486"  
repeat_region /note="match: GSS: Em:AQ225495"  
misc_feature /note="10383"  
repeat_region /note="MIR repeat: matches 79. .150 of consensus"  
misc_feature /note="match: GSS: Em:B14024"  
misc_feature /note="match: GSS: Em:B43656"  
repeat_region /note="MIR repeat: matches 20. .137 of consensus"  
repeat_region 12174..12445 /note="L2 repeat: matches 1988. .2275 of consensus"  
repeat_region 12444..12642 /note="MIR repeat: matches 63. .241 of consensus"  
misc_feature /note="match: STS: Em:G49301"  
repeat_region /note="MIR repeat: matches 174. .244 of consensus"  
repeat_region 13398..13698 /note="AluSp repeat: matches 1. .302 of consensus"  
repeat_region 13699..13810
```

```
repeat_region /note="MIR repeat: matches 76. .174 of consensus"  
repeat_region 13806..13919 /note="MIR repeat: matches 77. .189 of consensus"  
repeat_region 13945..14060 /note="MIR repeat: matches 24. .142 of consensus"  
repeat_region 14061..14367 /note="AluY repeat: matches 1. .301 of consensus"  
repeat_region 14368..14452 /note="MIR repeat: matches 141. .225 of consensus"  
repeat_region 14589..14679 /note="MIR repeat: matches 173. .262 of consensus"  
misc_feature 14597..15201 /note="match: GSS: Em:AQ553482"  
misc_feature 14616..15060 /note="match: GSS: Em:AQ370601"  
repeat_region 14868..15040 /note="MIR repeat: matches 49. .233 of consensus"  
repeat_region 15071..15188 /note="L2 repeat: matches 2112. .2239 of consensus"  
repeat_region 15304..15399 /note="MLTIB repeat: matches 1. .99 of consensus"  
repeat_region 15490..15662 /note="AluSg1 repeat: matches 2. .114 of consensus"  
repeat_region 15669..15727 /note="MLTIB repeat: matches 119. .178 of consensus"  
repeat_region 15728..16027 /note="AluSc repeat: matches 1. .299 of consensus"  
repeat_region 16028..16245 /note="MLTIB repeat: matches 178. .390 of consensus"  
repeat_region 16546..16854 /note="AluY repeat: matches 1. .300 of consensus"  
repeat_region 18296..18323 /note="MSTA repeat: matches 2. .29 of consensus"  
repeat_region 18324..18392 /note="MER66-internal repeat: matches 4919. .4993 of consensus"  
repeat_region 18393..18712 /note="AluJb repeat: matches 1. .311 of consensus"  
repeat_region 18713..19133 /note="MER66-internal repeat: matches 4548. .4919 of consensus"  
misc_feature complement(18872..19230)  
misc_feature /note="match: GSS: Em:AQ005063"  
repeat_region 19537..20290 /note="match: GSS: Em:B14179"  
repeat_region /note="HERVPH21 repeat: matches 4657. .5784 of consensus"  
repeat_region 20317..20382 /note="33 copies 2 mer ta 68 conserved"  
repeat_region 20513..20666 /note="77 copies 2 mer tt 70 conserved"  
repeat_region 20682..21008 /note="AluSg1 repeat: matches 1. .306 of consensus"  
repeat_region 21239..21553 /note="HUERS-P3 repeat: matches 4410. .4713 of consensus"  
repeat_region 21882..22254 /note="THEIB repeat: matches 1. .364 of consensus"  
repeat_region 22302..22537 /note="MER66-internal repeat: matches 2186. .2417 of consensus"  
repeat_region 22538..22850 /note="AluSp repeat: matches 1. .313 of consensus"  
repeat_region 22851..23801 /note="MER66-internal repeat: matches 1210. .2186 of consensus"  
repeat_region 23905..23989 /note="MER66-internal repeat: matches 3017. .3102 of consensus"  
Query Match 15.3%; Score 68.2; DB 9; Length 145880;  
Best Local Similarity 58.7%; Pred. No. 5.6e-09;  
Matches 118; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
```

```

Db 90879 AAGATCTGCATTGGTGAACACTGATCGGCCCTCCAGTGCCTTCTAGCCCATTTGT 90820
Qy 231 CTTACAGGGTCTCTACAACTGCCAGAGGATTTAAAGGAGGCCATGGGAGTCTTTGC 290
Db 90819 CTTCAAGGAAGCCCTTATATACATCACTTCATTCATTCATTCCTGCTGAGGATTCAC 90760
Qy 291 TTGCAGGTGCTCGGAATTAATACTACTCCAGCTTTGATGTTGAGCTGCATCATTCGAAGT 350
Db 90759 CTGTACCATCCCTAGTGTACCACTTTGGTTTGAATTTGAGCTGTTTCAGAGCTGTGC 90700
Qy 351 GAATATTTGGCTAATGAGAA 371
Db 90699 AAATGTGGGTATTATGAGAA 90679

RESULT 7
HS732E4
LOCUS
DEFINITION
Human DNA sequence from clone CTA-732E4 on chromosome 22q12.1
Contains ESTs, STSs and GSSs, complete sequence.
ACCESSION
AL008722
VERSION
AL008722.16 GI:5419644
KEYWORDS
HTG.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 90497)
Lloyd,D.
Direct Submission
Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 8, 1999 this sequence version replaced gi:5304865.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C/elegans/wormpep
from the human BAC library described in U.-J. Kim et al. (1996)
Genomics 34, 213-218.
VECTOR: pBelOBAcl1
This sequence is the entire insert of clone CTA-732E4 The true left
end of clone RP11-541J16 is at 5510 in this sequence. The true
right end of clone CTA-544A11 is at 41939 in this sequence.
FEATURES
source
1. 90497
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/map="q12.1"
/clone="CTA-732E4"
/clone.lib="CIT978SK-A2"
2. 631
repeat_region
/note="L1M1 repeat: matches 994. 1640 of consensus"
repeat_region
632. 757

```

```

/note="AluJb repeat: matches 1. 126 of consensus"
758. 1385
/note="L1P repeat: matches 3548. 4173 of consensus"
1374. 1869
/note="L1P repeat: matches 3221. 3716 of consensus"
1952. 2076
/note="L1MD2 repeat: matches 5142. 5259 of consensus"
2760. 3005
/note="L1MD2 repeat: matches 5259. 6173 of consensus"
3418. 3477
/note="30 copies 2 mer ac 90 conserved"
3423. 3478
/note="14 copies 4 mer caca 94 conserved"
3670. 3742
/note="MER82 repeat: matches 580. 653 of consensus"
3763. 4128
/note="MER82 repeat: matches 1. 386 of consensus"
3879. 4353
/note="match: GSS: Em:AQ559531"
4364. 4569
/note="L2 repeat: matches 2535. 2750 of consensus"
4612. 4930
/note="AluJo repeat: matches 1. 307 of consensus"
4934. 5220
/note="AluSx repeat: matches 1. 287 of consensus"
5221. 5268
/note="16 copies 3 mer taa 97 conserved"
complement(5266..5727)
/note="match: GSS: Em:AQ028870"
complement(5282..5740)
/note="match: GSS: Em:AQ557814"
5417. 5569
/note="MIR repeat: matches 2. 167 of consensus"
5645. 5708
/note="32 copies 2 mer at 71 conserved"
6469. 6726
/note="match: STS: Em:G03694; match: STS: Em:G03694"
7573. 7749
/note="MIR repeat: matches 22. 207 of consensus"
10322. 16495
/note="L1PA2 repeat: matches 7. 6146 of consensus"
16285. 16787
/note="match: GSS: Em:AQ186300"
16926. 16967
/note="21 copies 2 mer aa 76 conserved"
17850. 17971
/note="AluJo repeat: matches 19. 147 of consensus"
17972. 18285
/note="AluSg1 repeat: matches 1. 309 of consensus"
18286. 18472
/note="AluJo repeat: matches 147. 309 of consensus"
19710. 20005
/note="AluSx repeat: matches 1. 296 of consensus"
20789. 21262
/note="match: GSS: Em:B14443"
20811. 21058
/note="match: STS: Em:L04555"
21369. 21551
/note="MIR repeat: matches 2. 187 of consensus"
22494. 22697
/note="MIR repeat: matches 35. 236 of consensus"
23443. 23755
/note="AluSx repeat: matches 1. 312 of consensus"
24052. 24348
/note="AluSg1 repeat: matches 1. 295 of consensus"
25623. 25670
/note="24 copies 2 mer ac 77 conserved"
26003. 26192
/note="MIR repeat: matches 20. 216 of consensus"
27067. 27138
/note="Charlie4 repeat: matches 1902. 1961 of consensus"
27363. 27661
/note="AluYb8 repeat: matches 1. 306 of consensus"

```



Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (18-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Aug 18, 2002 this sequence version replaced gi:21699681.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L22077  
 Center clone name: 2384\_G\_12  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 1 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 135880: contig of 135880 bp in length.  
 \* Location/Qualifiers  
 1..135880  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="8"  
 /map="8"  
 /clone="CTD-2384G12"  
 /clone\_lib="CITDI Human BAC"  
 /clone\_a 25457 c 25972 g 42021 t  
 -----  
 BASE COUNT 42430 a 25457 c 25972 g 42021 t  
 ORIGIN  
 -----  
 Query Match 14.2%; Score 63.4; DB 2; Length 135880;  
 Best Local Similarity 86.4%; Pred. No. 1.5e-07;  
 Matches 70; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
 -----  
 Qy 365 TGAGAGCAAAATTTGGCTTAATAGGAAGAAATTTCTAAGCAGCAAGCATTCGAAGAGG 424  
 |||| || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 81101 TGAGAGAGATGATTCGGGTATTAGAGAAGAAATTTCTAAGCAGCAAGCATTCGAAGAGG 81042  
 -----  
 Qy 425 TGACTTGGTCTGCTTAAGG 445  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 81041 TGACTTGGTCTGCTTAAAG 81021  
 -----  
 RESULT 9  
 AP005265/c 51000 bp DNA linear PRI 07-JUN-2002  
 LOCUS Homo sapiens genomic DNA, chromosome 18, clone:RP11-78F17, complete  
 DEFINITION sequence.  
 ACCESSION AP005265  
 VERSION AP005265.2 GI:21328206  
 KEYWORDS HTG.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 -----  
 REFERENCE 1  
 Hattori, M., Toyoda, A., Taylor, T.D., Fujiyama, A., Yada, T.,  
 Totoki, Y., Watanabe, H. and Sakaki, Y.  
 Homo sapiens genomic DNA  
 Published Only in Database (2002)  
 2 (bases 1 to 51000)  
 Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,  
 Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
 Direct Submission  
 Submitted (24-MAY-2002) Masahira Hattori, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail: hattori@gsr.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,  
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
 On Jun 6, 2002 this sequence version replaced gi:21218138.  
 COMMENT  
 FEATURES  
 Location/Qualifiers  
 1..51000  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="18"  
 /map="18p"  
 /clone="RP11-78F17"  
 BASE COUNT 14110 a 9626 c 9391 g 17873 t  
 ORIGIN  
 -----  
 Query Match 14.1%; Score 63.2; DB 9; Length 51000;  
 Best Local Similarity 74.1%; Pred. No. 1.6e-07;  
 Matches 80; Conservative 0; Mismatches 28; Indels 0; Gaps 0;  
 -----  
 Qy 340 CATTGCAAGGTGAATATTGGCTAATGAGGAAGAAATTTGGCTTAATAAGGAAGAAATT 399  
 |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 44315 CCTAGAGATTTGTGAACCTTGAACCTGAAGAGAGATGATTTAGAGTATCAGGCAACAATT 44256  
 -----  
 Qy 400 TCTAAGCAGCAAGCAATTCACAGAGGTGACTTGGTGGCTGTTAAAGGCA 447  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 44255 TCTAAGCAGCAAGCAATTCACAGAGGTGACTTGGTGGCTGTTAAAGGCA 44208  
 -----  
 RESULT 10  
 AC027433 144514 bp DNA linear HTG 02-MAY-2000  
 LOCUS Homo sapiens chromosome 18 clone RP11-78F17 map 18, WORKING DRAFT  
 DEFINITION SEQUENCE, 36 unordered pieces.  
 AC027433  
 AC027433.2 GI:7677878  
 VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS Homo sapiens.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 -----  
 REFERENCE 1 (bases 1 to 144514)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,  
 Boguslavskiy, L., Boukhgaiter, B., Brown, A., Burkett, G.,  
 Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
 Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,  
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,  
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
 Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczyk, J.,  
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,  
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
 Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Spencer, B.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On May 2, 2000 this sequence version replaced gi:7342151.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

```
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7820
Center clone name: 78_F17
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 124489 bases at least Q40
Consensus quality: 133382 bases at least Q30
Consensus quality: 137772 bases at least Q20
Insert size: 139000; agarose-fp
Insert size: 141014; sum-of-contigs
Quality coverage: 3.5 in Q20 bases; agarose-fp
Quality coverage: 3.4 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1689: contig of 1689 bp in length
* 1690 1789: gap of 100 bp
* 1790 3019: contig of 1230 bp in length
* 3020 3119: gap of 100 bp
* 3120 4921: contig of 1802 bp in length
* 4922 5021: gap of 100 bp
* 5022 7426: contig of 2405 bp in length
* 7427 7526: gap of 100 bp
* 7527 9699: contig of 2173 bp in length
* 9700 9799: gap of 100 bp
* 9800 11899: contig of 2100 bp in length
* 11900 11999: gap of 100 bp
* 12000 14560: contig of 2561 bp in length
* 14561 14660: gap of 100 bp
* 14661 17033: contig of 2373 bp in length
* 17034 17133: gap of 100 bp
* 17134 18747: contig of 1614 bp in length
* 18748 18847: gap of 100 bp
* 18848 21078: contig of 2231 bp in length
* 21079 21178: gap of 100 bp
* 21179 23095: contig of 1917 bp in length
* 23096 23195: gap of 100 bp
* 23196 25510: contig of 2315 bp in length
* 25511 25610: gap of 100 bp
* 25611 27999: contig of 2389 bp in length
* 28000 28099: gap of 100 bp
* 28100 31388: contig of 3289 bp in length
* 31389 31488: gap of 100 bp
* 31489 34192: contig of 2704 bp in length
* 34193 34292: gap of 100 bp
* 34293 37246: contig of 2954 bp in length
* 37247 37346: gap of 100 bp
* 37347 39922: contig of 2576 bp in length
* 39923 40022: gap of 100 bp
* 40023 43695: contig of 3673 bp in length
* 43696 43795: gap of 100 bp
* 43796 48598: contig of 4803 bp in length
* 48599 48698: gap of 100 bp
* 48699 52567: contig of 3869 bp in length
* 52568 52667: gap of 100 bp
* 52668 55272: contig of 2605 bp in length
* 55273 55372: gap of 100 bp
* 55373 59485: contig of 4113 bp in length
* 59486 59585: gap of 100 bp
* 59586 63994: contig of 4409 bp in length
*
* 63995 64094: gap of 100 bp
* 64095 69097: contig of 5003 bp in length
* 69098 69197: gap of 100 bp
* 69198 73278: contig of 4081 bp in length
* 73279 73378: gap of 100 bp
* 73379 77729: contig of 4351 bp in length
* 77730 77829: gap of 100 bp
* 77830 82571: contig of 4742 bp in length
* 82572 82671: gap of 100 bp
* 82672 87792: contig of 5121 bp in length
* 87793 87892: gap of 100 bp
* 87893 92294: contig of 4402 bp in length
* 92295 92394: gap of 100 bp
* 92395 97160: contig of 4766 bp in length
* 97161 97260: gap of 100 bp
* 97261 102801: contig of 5541 bp in length
* 102802 102901: gap of 100 bp
* 102902 108682: contig of 5781 bp in length
* 108683 108782: gap of 100 bp
* 108783 115194: contig of 6412 bp in length
* 115195 115294: gap of 100 bp
* 115295 123411: contig of 8117 bp in length
* 123412 123511: gap of 100 bp
* 123512 132644: contig of 9133 bp in length
* 132645 132744: gap of 100 bp
* 132745 144514: contig of 11770 bp in length.
*
* FEATURES
* Location/Qualifiers
* 1..144514
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="18"
* /map="18"
* /clone="RP11-78F17"
* /clone_lib="RPCI-11 Human Male BAC"
*
* misc_feature
* 1..1689
* /note="assembly_fragment"
* 1790..3019
* /note="assembly_fragment"
* 3120..4921
* /note="assembly_fragment"
* 5022..7426
* /note="assembly_fragment"
* 7527..9699
* /note="assembly_fragment"
* 9800..11899
* /note="assembly_fragment"
* 12000..14560
* /note="assembly_fragment"
* 14661..17033
* /note="assembly_fragment"
* 17134..18747
* /note="assembly_fragment"
* 18848..21078
* /note="assembly_fragment"
* 21179..23095
* /note="assembly_fragment"
* clone_end:T7
* vector_side:right"
* 23196..25510
* /note="assembly_fragment"
* 25611..27999
* /note="assembly_fragment"
* 28100..31388
* /note="assembly_fragment"
* 31489..34192
* /note="assembly_fragment"
* 34293..37246
* /note="assembly_fragment"
* 37347..39922
* /note="assembly_fragment"
* 40023..43695
* /note="assembly_fragment"
* 43796..48598
* /note="assembly_fragment"
* 48699..52567
* /note="assembly_fragment"
* 52668..55272
* /note="assembly_fragment"
* 55373..59485
* /note="assembly_fragment"
* 59486..63994
* /note="assembly_fragment"
```

```
/note="assembly_fragment"
48699. .52567
/note="assembly_fragment"
52668. .55272
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
55373. .59485
/note="assembly_fragment"
59586. .63994
/note="assembly_fragment"
```

```
Query Match 14.1%; Score 63.2; DB 2; Length 144514;
Best Local Similarity 74.1%; Pred. No. 1.7e-07;
Matches 80; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 340 CATTGCAAGTCAATATTGGCTAATGAGGAAGCAAAATTTGGCTAATAAGCAAGAAATTT 399
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133169 CCTAGAGATTTGTGAACCTTGAACCTGGAAGAGATGATTTAGAGTATCAGGGAAGAAATTT 133228
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 400 TCTAAGCAGCAAGCATTCAAGAGGTGACTTGGGTGCTGTTAAAGCA 447
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133229 TCTAAGCAGCAAGCATTCAAGAGGTGACTTGGGTGCTGTTAAAGCA 133276
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 11
AC025112 149038 bp DNA linear HTG 08-APR-2000
LOCUS Homo sapiens chromosome 6 clone RP11-608N7 map 6, WORKING DRAFT
DEFINITION SEQUENCE, 27 unordered pieces.
ACCESSION AC025112
VERSION AC025112.2 GI:7528131
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bida,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Gadigan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lanazares,R., Landers,T., Lehoczyk,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stefan-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigglio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 8, 2000 this sequence version replaced gi:7158923.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
```

```
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6882
Center clone name: 608_N_7
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 132985 bases at least Q40
Consensus quality: 140748 bases at least Q30
Consensus quality: 143894 bases at least Q20
Insert size: 138000; agarose-fp
Insert size: 146438; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 3.4 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 165: contig of 165 bp in length
* 166 265: gap of 100 bp
* 266 1355: contig of 1050 bp in length
* 1356 1455: gap of 100 bp
* 1456 1752: contig of 297 bp in length
* 1753 1852: gap of 100 bp
* 1853 3336: contig of 1484 bp in length
* 3337 3436: gap of 100 bp
* 3437 4814: contig of 1378 bp in length
* 4815 4914: gap of 100 bp
* 4915 6853: contig of 1939 bp in length
* 6854 6953: gap of 100 bp
* 6954 9114: contig of 2161 bp in length
* 9115 9214: gap of 100 bp
* 9215 12071: contig of 2857 bp in length
* 12072 12171: gap of 100 bp
* 12172 14633: contig of 2462 bp in length
* 14634 14733: gap of 100 bp
* 14734 17354: contig of 2621 bp in length
* 17355 17454: gap of 100 bp
* 17455 21910: contig of 4456 bp in length
* 21911 22010: gap of 100 bp
* 22011 26287: contig of 4277 bp in length
* 26288 26387: gap of 100 bp
* 26388 31094: contig of 4707 bp in length
* 31095 31194: gap of 100 bp
* 31195 35408: contig of 4214 bp in length
* 35409 35508: gap of 100 bp
* 35509 40764: contig of 5256 bp in length
* 40765 40864: gap of 100 bp
* 40865 45639: contig of 4775 bp in length
* 45640 45739: gap of 100 bp
* 45740 50826: contig of 5087 bp in length
* 50827 50926: gap of 100 bp
* 50927 56617: contig of 5691 bp in length
* 56618 56717: gap of 100 bp
* 56718 63406: contig of 6689 bp in length
* 63407 63506: gap of 100 bp
* 63507 71321: contig of 7815 bp in length
* 71322 71421: gap of 100 bp
* 71422 81641: contig of 10220 bp in length
* 81642 81741: gap of 100 bp
* 81742 91982: contig of 10241 bp in length
* 91983 92082: gap of 100 bp
* 92083 102357: contig of 10275 bp in length
* 102358 102457: gap of 100 bp
* 102458 112952: contig of 10495 bp in length
* 112953 113052: gap of 100 bp
```

\* 113053 124237: contig of 11185 bp in length  
\* 124238 124337: gap of 100 bp  
\* 124338 135694: contig of 11357 bp in length  
\* 135695 135794: gap of 100 bp  
\* 135795 149038: contig of 13244 bp in length.

FEATURES

source  
1. .149038  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6"  
/clone="RP11-608N7"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. .165  
/note="assembly\_fragment  
clone\_end:T7  
vector\_side:right"  
misc\_feature  
266. .1355  
/note="assembly\_fragment"  
misc\_feature  
1456. .1752  
/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:left"  
misc\_feature  
1853. .3336  
/note="assembly\_fragment"  
misc\_feature  
3437. .4814  
/note="assembly\_fragment"  
misc\_feature  
4915. .6853  
/note="assembly\_fragment"  
misc\_feature  
6954. .9114  
/note="assembly\_fragment"  
misc\_feature  
9215. .12071  
/note="assembly\_fragment"  
misc\_feature  
12172. .14633  
/note="assembly\_fragment"  
misc\_feature  
14734. .17354  
/note="assembly\_fragment"  
misc\_feature  
17455. .21910  
/note="assembly\_fragment"  
misc\_feature  
22011. .26287  
/note="assembly\_fragment"  
misc\_feature  
26388. .31094  
/note="assembly\_fragment"  
misc\_feature  
31195. .35408  
/note="assembly\_fragment"  
misc\_feature  
35509. .40764  
/note="assembly\_fragment"  
misc\_feature  
40865. .45639  
/note="assembly\_fragment"  
misc\_feature  
45740. .50826  
/note="assembly\_fragment"  
misc\_feature  
50927. .56617  
/note="assembly\_fragment"  
misc\_feature  
56718. .63406  
/note="assembly\_fragment"  
misc\_feature  
63507. .71321  
/note="assembly\_fragment"  
misc\_feature  
71422. .81641  
/note="assembly\_fragment"  
misc\_feature  
81742. .91982  
/note="assembly\_fragment"  
misc\_feature  
92083. .102357  
/note="assembly\_fragment"  
misc\_feature  
102458. .112952  
/note="assembly\_fragment"  
misc\_feature  
113053. .124237  
/note="assembly\_fragment"  
misc\_feature  
124338. .135694  
/note="assembly\_fragment"  
misc\_feature  
135795. .149038  
/note="assembly\_fragment"  
BASE COUNT 44765 a 26643 c 27220 g 47805 t 2605 others  
ORIGIN

Query Match 14.1%; Score 63.2; DB 2; Length 149038;  
Best Local Similarity 74.1%; Pred. No. 1.7e-07;  
Matches 80; Conservative 0; Mismatches 28; Indels 0; Gaps 0;  
QY 340 CATTGCAAGTGAATATTGGCTAATGAGGAGCAAAATTTGGCTAATAAGGAGAAATT 399  
Db 90741 CCTAGAGATTTGTGAACCTTTGAACCTGAAAGAGGTATGATGATCAGGAGAGAAATT 90800  
QY 400 TCTAAGCAGCAAAAGCATTCAAGAGGTGACTTGGTGTCTGTTAAAGGCA 447  
Db 90801 TCTAAGCAGCAAAAGCATTCAAGAGGTGACTTGGTGTCTGTTAAAGGCA 90848

RESULT 12  
AC015553/c  
LOCUS  
DEFINITION  
Homo sapiens 9 BAC RP11-100N10 (Roswell Park Cancer Institute Human BAC Library) complete sequence.  
AC015553  
AC015553.21 GI:13489133  
HTG.  
Homo sapiens.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 162063)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,B., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., and Gibbs,R.  
Direct Submission  
Unpublished  
2 (bases 1 to 162063)  
Worley,K.C.  
Direct Submission  
Submitted (17-NOV-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

REFERENCE  
AUTHORS

Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 162063)  
Worley, K.C.

TITLE  
JOURNAL

Direct Submission  
Submitted (30-MAR-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
AUTHORS

4 (bases 1 to 162063)  
Worley, K.C.

TITLE  
JOURNAL

Direct Submission  
Submitted (12-JUL-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
AUTHORS

5 (bases 1 to 162063)  
Worley, K.C.

TITLE  
JOURNAL

Direct Submission  
Submitted (13-JUL-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

## COMMENT

On Mar 30, 2001 this sequence version replaced gi:13487890.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
[gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.  
Overlapping clones are noted at the beginning and end of the  
Features listing.

## ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.  
Reports of lowest quality individual bases and measures of base  
quality are listed below. Description of the metrics can be found  
at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

## QUALSTAT-REPORT-----

```
----- Summary Statistics -----
Contig length: 162063
Phrap values in estimate: 161326
Average error rate (BCM-Phrap estimate): 9.86316e-06
Fraction of Phrap values less than 40 : 0.00234308
Number of consensus changing edits: 18
Number of N's in consensus : 0
```

```
----- Consensus changing edits -----
Position Original+Context Edited+Context
10095 caccacatt(t)acctccaaa
10096 accacacatt(a)ccctccaaa
```

37628

41530

68252

68273

68274

68288

87792

98412

98413

113818

113821

113849

122621

133372

134847

160961

```
aggaaggaag(n)aaggaaggaag
ccagcaataa(a)caagtgagaat
gctggaattg(n)tccttatttt
catctaggtg(n)nggttccatg
atctaggtg(c)tggttccatg
tccattttg(n)atacatagat
tgggtataag(n)caacagaaga
acgttaacta(n)ccaagggtccc
cgtaacttan(n)caaggtcccc
tgatactgaa(n)ccnctaaaaa
tactgaancc(n)ctaataaaaa
aaaaaanaa(n)gactaataa
atggaccatg(n)tgctttgtga
gtttataaaa(n)caataactta
tcctttactg(n)ggccctatta
aaaaaaaat(c)caaaaacttg
```

----- Distribution of Quality &lt; 40 Bases -----

```
5001
4501
4001
3501
3001
2501
2001
1501
1001
501
01
# bases
Phrap Value Range
5 10 15 20 25 30 35 40
```

Version: 1.01 qxfo.

## FEATURES

## source

```
1. 162063
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-100N10"
706..734
/repeat_family="AT-rich"
complement(738..930)
/repeat_family="L1MA4"
complement(943..1086)
/repeat_family="L1ME3A"
1979..2255
/repeat_family="L1MA8"
2560..2722
/repeat_family="AluJb"
2723..2831
/repeat_family="GA-rich"
2832..2859
/repeat_family="(GAAA)n"
2871..3150
/repeat_family="AluY"
3151..3335
/repeat_family="(GAAA)n"
3336..3346
/repeat_family="AluY"
3347..3481
/repeat_family="L1M4"
3482..4108
/repeat_family="L1MA4"
4109..4207
/repeat_family="L1M4"
```

Query Match

Best Local Similarity

14.0%; Score 62.6; DB 9; Length 162063;

79.6%; Pred. No. 2.6e-07;

```
Matches 74; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 355 ATTGGCTAATGAGGAACAATTTTGGCTAATAAGCAAGAATTTCTAAGCAGCAAAAGC 414
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29373 ACTTGAACCTGAGAGAGATGATTAGGTATGGCAGAGAAATTTCTAAGCAGCAAAAGC 29314
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 415 ATTTCAAGAGTGACTTGGTGGTGCTGTTAAAGGCA 447
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29313 ATTTCAAGAGTGACTTGGTGGTGCTGTTAAAGACA 29281
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
AL161623/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AL161623
Homo sapiens chromosome 9 clone RP11-1K21, *** SEQUENCING IN
PROGRESS ***, 20 unordered pieces.
AL161623
AL161623.12 GI:9930848
HTG: HTGS-PHASE1; HTGS-CANCELLED.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 175134)
McLay, K.
Direct Submission
Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 27, 2000 this sequence version replaced gi:9863606.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bA1K21
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator ET-amersham; 24% of reads Chemistry:
Dye-terminator Big Dye; 75% of reads
Consensus quality: 164370 bases at least Q40
Consensus quality: 168213 bases at least Q30
Consensus quality: 170527 bases at least Q20
Insert size: 173234; sum-of-contigs
Insert size: 168172; 4.4% error; agarose-fp
Quality coverage: 4.13x in Q20 bases; sum-of-contigs Quality
coverage: 4.57x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 3511: contig of 3511 bp in length
* 3512 3611: gap of 100 bp
* 3612 25895: contig of 22284 bp in length
* 25896 25995: gap of 100 bp
* 25996 31543: contig of 5548 bp in length
* 31544 31643: gap of 100 bp
* 31644 39248: contig of 7605 bp in length
* 39249 39348: gap of 100 bp
* 39349 49948: contig of 10600 bp in length
* 49949 50048: gap of 100 bp
* 50049 66131: contig of 16083 bp in length
* 66132 66231: gap of 100 bp
* 66232 80131: contig of 13900 bp in length
* 80132 80231: gap of 100 bp
* 80232 89785: contig of 9554 bp in length
*
89786 89885: gap of 100 bp
89886 92978: contig of 3093 bp in length
92979 93078: gap of 100 bp
93079 97528: contig of 4450 bp in length
97529 97628: gap of 100 bp
97629 101520: contig of 3892 bp in length
101521 101620: gap of 100 bp
101621 112966: contig of 11346 bp in length
112967 113066: gap of 100 bp
113067 119217: contig of 6151 bp in length
119218 119317: gap of 100 bp
119318 121556: contig of 2239 bp in length
121557 121656: gap of 100 bp
121657 134400: contig of 12744 bp in length
134401 134500: gap of 100 bp
134501 144934: contig of 10434 bp in length
144935 145034: gap of 100 bp
145035 148818: contig of 3784 bp in length
148819 148918: gap of 100 bp
148919 152178: contig of 3260 bp in length
152179 152278: gap of 100 bp
152279 165020: contig of 12742 bp in length
165021 165120: gap of 100 bp
165121 175134: contig of 10014 bp in length.
Location/Qualifiers
1..175134
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-1K21"
/clone_lib="RPCI-11.1"
1..3511
/note="assembly_fragment:00168
fragment_chain:1"
misc_feature
3612..25895
/note="assembly_fragment:00194
fragment_chain:1"
misc_feature
25996..31543
/note="assembly_fragment:00852
fragment_chain:2"
misc_feature
31644..39248
/note="assembly_fragment:00602
fragment_chain:2"
misc_feature
39349..49948
/note="assembly_fragment:01593
fragment_chain:3"
misc_feature
50049..66131
/note="assembly_fragment:01966
fragment_chain:3"
misc_feature
66232..80131
/note="assembly_fragment:02033
fragment_chain:4"
misc_feature
80232..89785
/note="assembly_fragment:00372
fragment_chain:4"
misc_feature
89886..92978
/note="assembly_fragment:02305
fragment_chain:5"
misc_feature
93079..97528
/note="assembly_fragment:00011
fragment_chain:5"
misc_feature
97629..101520
/note="assembly_fragment:00955"
misc_feature
101621..112966
/note="assembly_fragment:01248"
misc_feature
113067..119217
/note="assembly_fragment:01397"
misc_feature
119318..121556
/note="assembly_fragment:01523"
misc_feature
121657..134400
/note="assembly_fragment:01568"
misc_feature
134501..144934
/note="assembly_fragment:01617"
```

```
misc_feature 145035..148818
/note="assembly_fragment:01696"
misc_feature 148919..152178
/note="assembly_fragment:01972"
misc_feature 152279..165020
/note="assembly_fragment:00951
fragment_chain:6"
misc_feature 165121..175134
/note="assembly_fragment:01119
fragment_chain:6
clone_end:T7
vector_side:right"
BASE COUNT 51517 a 36761 c 35231 g 49719 t 1906 others
ORIGIN
...
Query Match 14.0%; Score 62.6; DB 2; Length 175134;
Best Local Similarity 79.6%; Pred. No. 2.7e-07;
Matches 74; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 355 ATTTGGCTAATGAGGACAAATTTTGGCTAATAAGGAAGAAATTTCTTAAGCAGCAAAAGC 414
||||| ||||| || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83975 ACTTTGAACCTTGAGAGAGATGATTTAGGTATGGCAGAGAATTTCTTAAGCAGCAAAAGC 83916
QY 415 ATTTCAAGAGGTCACCTTGGTGCTGTTAAAGGCA 447
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83915 ATTTCAAGAGGTCACCTTGGTGCTGTTAAAGGCA 83883

RESULT 14
AL157884 176932 bp DNA linear PRI 13-SEP-2001
LOCUS Human DNA sequence from clone RP11-462B18 on chromosome
9p13.1-21.1, complete sequence.
ACCESSION AL157884
VERSION AL157884.9 GI:15029097
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 176932)
Direct Submission
Submitted (13-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerquest@sanger.ac.uk
On Jul 29, 2001 this sequence version replaced gi:15026899.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; SW,
SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
RP11-462B18 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACg3.6

This sequence is the entire insert of clone RP11-462B18 The true
right end of clone RP11-555J4 is at 30638 in this sequence. The
true right end of clone RP11-1K21 is at 97783 in this sequence.

FEATURES
source
1..176932
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="p13.1-21.1"
/clone="RP11-462B18"
/clone_lib="RPCI-11.2"
/6326..77014
/note="CpG island"
/evidence=not_experimental
BASE COUNT 54421 a 37529 c 35943 g 49039 t
ORIGIN
...
Query Match 14.0%; Score 62.6; DB 9; Length 176932;
Best Local Similarity 79.6%; Pred. No. 2.7e-07;
Matches 74; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 355 ATTTGGCTAATGAGGACAAATTTTGGCTAATAAGGAAGAAATTTCTTAAGCAGCAAAAGC 414
||||| ||||| || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 89950 ACTTTGAACCTTGAGAGAGATGATTTAGGTATGGCAGAGAATTTCTTAAGCAGCAAAAGC 89891
QY 415 ATTTCAAGAGGTCACCTTGGTGCTGTTAAAGGCA 447
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 89890 ATTTCAAGAGGTCACCTTGGTGCTGTTAAAGGCA 89858

RESULT 15
TMSHP27
LOCUS Tamas asiaticus mRNA for HP-27, complete cds.
DEFINITION TMSHP27
ACCESSION D12976
VERSION D12976.1 GI:287471
KEYWORDS HP-27; collagen-like domain; hibernation-related protein; plasma
protein.
SOURCE Tamas asiaticus liver cDNA to mRNA, clone:PCM27-3.
ORGANISM Tamas sibiricus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Scluridae; Sclurinae;
Tamas.
1 (bases 1 to 1385)
AUTHORS Takamatsu,N., Ohba,K., Kondo,J., Kondo,N. and Shiba,T.
TITLE Hibernation-associated gene regulation of plasma proteins with a
collagen-like domain in mammalian hibernators
JOURNAL Mol. Cell. Biol. 13 (3), 1516-1521 (1993)
MEDLINE 93180798
REFERENCE 2 (bases 1 to 1385)
AUTHORS Takamatsu,N.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1992) Nobuhiko Takamatsu, School of Hygienic
Sciences, Kitasato University, Laboratory of Molecular Biology;
1-15-1 Kitasato, Sagamihara, Kanagawa 228, Japan (Tel:0427-78-9408,
Fax:0427-78-9403)
Location/Qualifiers
1..1385
/organism="Tamas sibiricus"
/db_xref="taxon:64680"
/clone="PCM27-3"
/tissue_type="liver"
65..712
/note="collagen-like domain at nt 191-307"
/codon_start=1
/product="HP-27"
/protein_id="BAA02353.1"
/db_xref="GI:287472"
/translation="MYEAGKRASFMGAGIWTALSLVLMHVVCSETQGNPSCNVPVPGP
OGPGMRGPGCTGKPGPGKNGNPGLPGLPCPGMGTVNCNHSKGTSAFAVKANELPPA
PSQPVIFREALHDAQHDFLATGVTCTCPVGLYQFGFHEAVQRAVKYSLMRNGTQVM
EREAQDQGYEHISGTAILQLGMEDRVWLENKLSOTDLERGTQVAVFSGFLIHN"
65..154
sig_peptide
```

```
mat_peptide 155..709
/product="HP-27"
polyA_signal 1369..1374
polyA_site 1385
BASE COUNT 439 a 285 c 311 g 350 t
ORIGIN
Query Match 14.0%; Score 62.4; DB 10; Length 1385;
Best Local Similarity 55.6%; Pred. No. 2e-07;
Matches 120; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
QY 156 CTCTCCTGATATAGAAAGCTCAGCCCTTTACTGTGAAGCTCAGTGGGAAAACCTTCCTCTCC 215
Db ||| || || || || || || || || || || || || || || || || || || || || ||
313 CTGCCACAGCAAGGAACATCGCCCTTTGAGTGAAGGCAAAATGAGCTGCCCCCAGCTCC 372
QY 216 TTTCAAGCCCATCATCTTCACAGGGTCTGTACAATGCCAGAGGGATTAAAGGAGGC 275
Db || || || || || || || || || || || || || || || || || || || || || ||
373 CTCCCAGCCCGTGAUCTTCAAGGAAGCCCTGCATGACGCTCAGGGACACTTTGATCTGGC 432
QY 276 CATGGGAGTCTTTGCTTGCAAGGTGCCTGGGAATTACTCCAGCTTTGATGTTGAGCT 335
Db || || || || || || || || || || || || || || || || || || || || || ||
433 CACTGTGTGTTTCACCTGCCAGTCCCGAGACTCTTACCAGTTTGGATTTCACATTGAAGC 492
QY 336 GCATCATTTGCAAGGTGAATATTTGGCTAATGAGGAA 371
Db || || || || || || || || || || || || || || || || || || || || || ||
493 TGTCCAGAGGGGTGTGAAGGTGAGCCTCATGAGAA 528

Search completed: February 20, 2003, 05:05:34
Job time : 2438.16 secs
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:48:42 ; Search time 391.859 Seconds  
(without alignments)  
6406.126 Million cell updates/sec

Title: US-09-997-610-1-copy\_2\_156

Perfect score: 155

Sequence: 1 atagtggtacatacctgtctt.....agtgaatgacaaaatgcc 155

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: gb\_gss:\*

18: em\_gss\_hum:\*

19: em\_gss\_inv:\*

20: em\_gss\_pln:\*

21: em\_gss\_vrt:\*

22: em\_gss\_fun:\*

23: em\_gss\_mam:\*

24: em\_gss\_mus:\*

25: em\_gss\_other:\*

26: em\_gss\_pro:\*

27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39.6	25.5	342	14	244751 HSC28C121 n
2	39.4	25.4	638	9	AU205784 AU205784
3	38.6	24.9	483	13	BI322379 kx20d10.y
4	38.6	24.9	531	10	BE121287 UI-R-CAO-
5	38.6	24.9	552	13	BI863711 kx47e02.y
6	37.8	24.4	352	9	AI617644 zehn1908.

7	37.8	24.4	770	13	BJ121588
8	37.4	24.1	468	10	BE580165
9	37.4	24.1	889	14	BQ939199
c 10	37.2	24.0	651	17	CNS05LPT
c 11	37.2	24.0	1101	17	CNS04XPG
12	37	23.9	454	10	BE581564
13	36.8	23.7	525	10	AW675997
14	36.8	23.7	575	10	AW179710
15	36.8	23.7	585	10	AW179840
16	36.8	23.7	605	10	AW347957
17	36.8	23.7	624	12	BE758489
18	36.8	23.7	640	10	AW179745
19	36.8	23.7	647	10	AW409404
20	36.8	23.7	647	10	AW409405
21	36.8	23.7	648	10	AW874708
22	36.8	23.7	654	10	AW348036
23	36.8	23.7	695	10	BE285930
24	36.8	23.7	701	10	AW179849
25	36.8	23.7	776	13	BI554667
26	36.8	23.7	801	9	AA585675
27	36.2	23.4	360	14	D69730
28	36.2	23.4	407	10	BE581103
29	36.2	23.4	420	10	BE579707
30	36.2	23.4	441	10	BE581944
31	36.2	23.4	478	10	BE579460
32	36.2	23.4	491	10	BE579541
33	36.2	23.4	492	10	BE579271
34	36.2	23.4	495	10	BE579572
35	36.2	23.4	632	13	BJ136441
c 36	36.2	23.4	726	13	BJ108893
c 37	36	23.2	450	13	BJ002073
38	36	23.2	461	13	BI323061
39	36	23.2	865	14	BQ721294
40	35.8	23.1	407	13	BI742359
41	35.8	23.1	458	14	BQ35482
42	35.8	23.1	467	13	BI323796
43	35.8	23.1	483	10	BE579244
44	35.8	23.1	675	12	BG310518
45	35.8	23.1	786	10	BE383449

#### ALIGNMENTS

RESULT 1

244751

LOCUS

DEFINITION HSC28C121 normalized infant brain cDNA Homo sapiens linear EST 14-NOV-1994

ACCESSION C-28C12, mRNA sequence.

VERSION 244751

KEYWORDS 244751.1 GI:573911

SOURCE EST.

ORGANISM human.

REFERENCE Homo sapiens

AUTHORS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 342)

Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes ,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y., Sebastiani-Kabaktchis,C. and Tessier,A.

IMAGE: molecular integration of the analysis of the human genome and its expression

C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)

95277534

CONTACT: Genethon

Genexpress-Genethon

Genethon Centre de recherche sur le Genome Humain

1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE

Tel: 33169472800

Fax: 33160778698

Email: genexpress@genethon.fr

Single read.

```
Genexpress_library_id: C; Genexpress_sequence_id: ylc-28cl2
Seq primer: (-21)JM13_universal.
Location/Qualifiers
  1. 342
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone_lib="c-28cl2"
  /clone_lib="normalized infant brain cDNA"
  /sex="Female"
  /tissue_type="total brain"
  /dev_stage="3 months old"
  /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
  Site_2: NotI; sex:Female; dev_stage=3 months old;
  isolate=muscular atrophy patient; tissue_type=total brain
  ; total mRNA was oligo-(dT) primed and directionally
  cloned 5' -> 3' into the HindIII -> NotI sites of the
  lafmid BA vector. Clone library from B.Souares, Psychiatry
  Dept. Columbia University, USA. Normalization_method:
  Bento Soares, P.N.A.S in press"
BASE COUNT      78 a 79 c 89 g 88 t
ORIGIN
Query Match      25.5%; Score 39.6; DB 14; Length 342;
Best Local Similarity 55.1%; Pred. No. 1;
Matches 75; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
QY 13 CCGTGTCTTAATAACGGCAGTCATTGACGATGTAGAGTTCGTGGACCTCCAGCACACCC 72
Db 48 CCGGTCCGACTCTCTCGCGGTGTGGTCAGTGTGAATTTGTGACAGCTGCAGTTGCTCC 107
QY 73 AGGCCCCCAGAGAAGTGGGGCTCTCCTGGTGCACCAGGTTTACCACAATATACAGGAGAA 132
Db 108 GCCGCCCGAGCAGCGAGGAGTCTACCGTGGCTCCNGANTTTCCCAAAATTCAGCAGCA 167
QY 133 ATAAAGTGAATGACAA 148
Db 168 GAAATTCAGTGACTA 183

RESULT 2
AU205784
LOCUS AU205784 638 bp mRNA linear EST 17-JUL-2001
DEFINITION AU205784 unpublished oligo-capped cDNA library, stage L1
Caenorhabditis elegans cDNA clone yk855b02 5', mRNA sequence.
ACCESSION AU205784
VERSION AU205784.1 GI:14838413
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 638)
AUTHORS Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE A complementary view of the C.elegans genome
JOURNAL Unpublished (2001)
COMMENT Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
FEATURES
  source
    1..638
    Location/Qualifiers
      /organism="Caenorhabditis elegans"
      /strain="N2"
      /db_xref="taxon:6239"
      /clone_lib="yk855b02"
      /clone_lib="unpublished oligo-capped cDNA library, stage
      L1"
      /sex="Hermaphrodite"
      /tissue_type="whole animal"
```

```
/dev_stage="L1"
/note="The AD-wrmcDNA library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into pPC86"
BASE COUNT      193 a 176 c 190 g 79 t
ORIGIN
Query Match      25.4%; Score 39.4; DB 9; Length 638;
Best Local Similarity 56.6%; Pred. No. 1.2;
Matches 73; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 21 AATAACGGCAGTCATTAGCATGTAGAGTTCGTGGACCTCCAGCACACCCAGGCCCC 80
Db 504 AATGAAGGGAGACTCTGGCTCTCCAGGACCACAGGACTCCAGGACACCCAGGTGTCC 563
QY 81 AGAAGAAGTGGGCCCTCCTGGTGCACCAGGTTTACCACAATATACAGGAGAAATAGTGA 140
Db 564 AGGAGACAAGGAGTTCGGAGGTGTCCAGGATTACCGAGATTCCAGGACCAAGGGAGA 623
QY 141 AATGACAAA 149
Db 624 TCTCGGAAA 632

RESULT 3
BI322379
LOCUS BI322379 483 bp mRNA linear EST 30-JUL-2001
DEFINITION kx20d10.y3 Parastrongyloides trichosuri FL pAMP1 v1 Chiapelli
McCarte Parastrongyloides trichosuri cDNA 5' similar to TR:Q61436
Q61436 PROCOLLAGEN, TYPE IV, ALPHA 5 ;contains element PTR5
repetitive element ;, mRNA sequence.
ACCESSION BI322379
VERSION BI322379.1 GI:15001565
KEYWORDS EST.
SOURCE Parastrongyloides trichosuri.
ORGANISM Parastrongyloides trichosuri
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Panagrolaimoidea; Strongyloidea; Parastrongyloides.
REFERENCE 1 (bases 1 to 483)
AUTHORS McCarte,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.,
Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarisshvili,R.,
Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,
Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
Wilson,R.
TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: McCarte JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Brandi Chiapelli and Dr. James
McCarte (bchiapel@watson.wustl.edu & jmccarte@watson.wustl.edu) at
Washington University, St. Louis. DNA Sequencing by: Washington
University Genome Sequencing Center St. Louis.
Seq primer: -40RP from Gibco
High quality sequence stop: 428.
FEATURES
  source
    1..483
    Location/Qualifiers
      /organism="Parastrongyloides trichosuri"
      /db_xref="taxon:131310"
      /clone_lib="Parastrongyloides trichosuri FL pAMP1 v1
      Chiapelli McCarte"
      /dev_stage="Free Living"
      /lab_host="DH10B"
```

James McCarter at Washington University, St. Louis. The cDNA was made by using Dynabead oligo-dT priming (Dyna). PCR based library using a modified protocol from the SMART PCR cDNA Synthesis Kit from Clontech. Directionally cloned into the UDG sites of pAMP1. Nematodes were provided by Dr. Warwick Grant of AgResearch, New Zealand (warwick.grant@agresearch.co.nz).

BASE COUNT 157 a 103 c 123 g 100 t

Query Match 24.9%; Score 38.6; DB 13; Length 483;  
Best Local Similarity 61.4%; Pred. No. 1.9;  
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 45 AGAAGTTGCTGACCTCCAGCACACCCAGGCCGCCAGAGAAGTGGGCCCTCTGGTGC 104

Db 237 AGGAAAGATGGATTACCAAGATTCCAGGAATTAAGAGGAAGTGGTTCCCTGGAG 296

Qy 105 ACCAGGTTTACCAATATACAGGAGAAATAGTGAATGA 145

Db 297 ACCAGGTATGCCAGGATTGAAAGGAGAAATTAGAGAATTTA 337

#### RESULT 4

BE121287/c

LOCUS

DEFINITION BE121287 531 bp mRNA linear EST 13-JUN-2000  
UI-R-CA0-baw-b-08-0-UI.s1 UI-R-CA0 Rattus norvegicus cDNA clone  
UI-R-CA0-baw-b-08-0-UI 3', mRNA sequence.

ACCESSION BE121287

VERSION BE121287.1

KEYWORDS GI:8513392

SOURCE EST.

ORGANISM Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 531)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dT track served to identify it as a clone from the

normalized corpus-striatum library cDNA Library Preparation: M.B.

Soares Lab Clone distribution: clones will be available through

Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA-Yes.

Location/Qualifiers

1. .531

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db.xref="taxon:10116"

/clone="UI-R-CA0-baw-b-08-0-UI"

/clone\_lib="UI-R-CA0"

/lab\_host="DH10B (Life Technologies)"

/note="vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site.1: Not I; Site.2: Eco RI; The UI-R-CA0

library is a subtracted library derived from the following

tissues: thalamus, cerebellum, hypothalamus, medulla, pons

, midbrain, cerebral cortex, corpus striatum, testis, and

hippocampus. For a detailed description of the library

from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)

TAG\_LIB=UI-R-CA0

TAG\_TISSUE=corpus-striatum

TAG\_SEQ=CTAGG

BASE COUNT 124 a 134 c 128 g 145 t

ORIGIN

Query Match 24.9%; Score 38.6; DB 10; Length 531;

Best Local Similarity 59.6%; Pred. No. 1.9;

Matches 65; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 40 CATGTAGAGTTTGGCTCCAGCACACCCAGGCCGCCAGAGAAGTGGGCCCTCT 99

Db 526 CTTCTTGGCCCTCACAGGACAGTCAGGGCTCCCTGCCCTGGACAGCGGACTCCT 467

Qy 100 GGTGCACCAAGTTTACCACACATATACAGGAGAAATAGTGAATGACAA 148

Db 466 GGAGTTCAGGGTTCCAGGTTTCTAAGGTGAATGGCGTCATGGGAA 418

#### RESULT 5

BI863711

LOCUS

DEFINITION BI863711 552 bp mRNA linear EST 10-OCT-2001  
kx47e02.y1 Parastrongyloides trichosuri FL pAMP1 v1 Chiapelli  
McCarter Parastrongyloides trichosuri cDNA 5', similar to  
SW:CA24\_ASCSU P27393 COLLAGEN ALPHA 2(IV) CHAIN PRECURSOR. [1]  
;contains element PTR5 repetitive element ;, mRNA sequence.

ACCESSION BI863711

VERSION BI863711.1

KEYWORDS GI:16005501

SOURCE EST.

ORGANISM Parastrongyloides trichosuri.

Parastrongyloides trichosuri

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;

Panagrolaimoidea; Strongyloidea; Parastrongyloidea.

1 (bases 1 to 552)

McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.

, Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.

, Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagareishvili, R.

, Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe

, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.

, Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and

Wilson, R.

The Washington Univ. Nematode EST Project, 1999

Unpublished (1999)

Contact: McCarter JP

The Washington Univ. Nematode EST Project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

The library was constructed by Brandi Chiapelli and Dr. James

McCarter (bchiapell@watson.wustl.edu & jmcarter@watson.wustl.edu) at

Washington University, St. Louis. DNA Sequencing by: Washington

University Genome Sequencing Center St. Louis.

Seq primer: -40RP from Gibco

High quality sequence stop: 395.

Location/Qualifiers

1. .552

/organism="Parastrongyloides trichosuri"

/db.xref="taxon:131310"

/clone\_lib="Parastrongyloides trichosuri FL pAMP1 v1

Chiapelli McCarter"

/dev\_host="Free Living"

/lab\_host="DH10B"

/note="vector: pAMP1 (Gibco); Site.1: NotI; Site.2: SalI;

The library was constructed by Brandi Chiapelli and Dr.

James McCarter at Washington University, St. Louis. The

cDNA was made by using Dynabead oligo-dT priming (Dyna).

PCR based library using a modified protocol from the

#### FEATURES

source



```

ACCESSION      BE580165
VERSION        BE580165.1  GI:9831107
KEYWORDS
SOURCE
ORGANISM       Strongyloides stercoralis.
REFERENCE
AUTHORS        McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,
               Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,
               Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagareishvili, R.,
               Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe
               , M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,
               , Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and
               Wilson, R.
TITLE          The Washington Univ. Nematode EST Project, 1999
JOURNAL        Unpublished (1999)
COMMENT        Contact: McCarter JP
               The Washington Univ. Nematode EST Project, 1999
               Washington University School of Medicine
               4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
               Tel: 314 286 1800
               Fax: 314 286 1810
               Email: est@watson.wustl.edu
               The library was constructed by Dr. Thomas Nutman and colleagues of
               NIAID, NIH (ncutman@nih.gov). DNA Sequencing by: Washington
               University Genome Sequencing Center St. Louis.
               High quality sequence stop: 325.
               Location/Qualifiers
FEATURES       source
               1..468
               /organism="Strongyloides stercoralis"
               /strain="Rhabditiform larvae obtained from gerbills"
               /db_xref="taxon:6248"
               /clone_lib="TBN95TM-SSR"
               /lab_host="XL-1 Blue MRP" (Stratagene)
               /note="Vector: Lambda Uni-ZAP XR (Stratagene); Site_1:
               EcoRI; Site_2: XhoI; mRNA was purified from 2 x 10B3
               rhabditiform larvae which had been isolated from gerbills
               experimentally infected with larvae originally isolated
               from experimentally infected dogs. cDNA was constructed
               and, using adaptors, was cloned unidirectionally into the
               vector from the EcoRI site to the XhoI site. The library
               has an unamplified titer of 1 x 10E5 pfu/ml and an
               amplified, undiluted titer of 9 x 10E11 pfu/ml. The
               average insert size of the unamplified library is 675 bp
               (range, 100-1700)."
BASE COUNT     157 a 92 c 134 g 85 t
ORIGIN
               Query Match      24.1%; Score 37.4; DB 10; Length 468;
               Best Local Similarity 57.1%; Pred. No. 4;
               Matches 68; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy  24  AACGCCATGTCATGAGCATGTAGAGTGTCTGGACCTCAGCACACCCAGGCCCCAGCA 83
      || || || || || || || || || || || || || || || || || || || ||
Db  126  AAGAGGAGAAAAGGACAAAAGGAGAACAGGACACAGGACCAAGATGGATCTCCAGG 185
      || || || || || || || || || || || || || || || || || || || ||
Qy  84  AGAAGTGGGGCTCTCGTGCACAGGTTTACCACAATATACAGGAGAAATAAGTGAAA 142
      || || || || || || || || || || || || || || || || || || || ||
Db  186  AGACAGGACCAAGAGGTGAACAAGGTTTACAGGAGATATTGGACCATCAGGAGAAA 244
      || || || || || || || || || || || || || || || || || || || ||

RESULT 9
B0939199
LOCUS          B0939199      889 bp      mRNA      linear      EST 21-AUG-2002
DEFINITION     AGNCOURT_8922132 NCI_CGAP_Co24 Mus musculus cDNA clone
               IMAGE:6395968 5', mRNA sequence.
ACCESSION      B0939199
VERSION        B0939199.1  GI:22354677
KEYWORDS
SOURCE          house mouse.
ORGANISM        Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 889)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL        Unpublished (1999)
COMMENT        Contact: Robert Strausberg, Ph.D.
               Email: c9apbs-r@mail.nih.gov
               Tissue procurement: The Cepko Laboratory
               cDNA Library Preparation: Life Technologies, Inc.
               DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Agencourt Bioscience Corporation
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLAM13891 row: h column: 17
               High quality sequence stop: 672.
               Location/Qualifiers
FEATURES       source
               1..889
               /organism="Mus musculus"
               /strain="FVB/N"
               /db_xref="taxon:10090"
               /clone="IMAGE:6395968"
               /clone_lib="NCI_CGAP_Co24"
               /lab_host="DH10B (T1 phage-resistant)"
               /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
               Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
               Average insert size 1.6 kb. Constructed by Life
               Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT     202 a 251 c 292 g 140 t
ORIGIN
               Query Match      24.1%; Score 37.4; DB 14; Length 889;
               Best Local Similarity 61.5%; Pred. No. 4.1;
               Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy  53  CTGGACCTCCAGCACACCCAGGCCGCCAGAGAGAGTGGGGCTCTGTGTGCACAGGTT 112
      | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  600  CAGGACAGTCAGCGCTCCCTGGCCCTTCTCTGGACAGCAGGGGACACCTGGAGTTCAGG 659
      || || || || || || || || || || || || || || || || || || || ||
Qy  113  TACCAACATATACAGGAGAAATAAGTGAATGACAA 148
      || || || || || || || || || || || || || || || || || || || ||
Db  660  TCCAGGTCTTANAGTGAATGGGTGTCATGGGAA 695
      || || || || || || || || || || || || || || || || || || || ||

RESULT 10
CNS05LPT/C
LOCUS          CNS05LPT      651 bp      DNA      linear      GSS 26-MAY-2000
DEFINITION     Tetraodon nigroviridis genome survey sequence T7 end of clone
               048C07 of library A from Tetraodon nigroviridis, genomic survey
               sequence.
ACCESSION      AL343082
VERSION        AL343082.1  GI:8236840
KEYWORDS        GSS: genome survey sequence.
SOURCE          Tetraodon nigroviridis.
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
               Acanthopterygii; Acanthopterygii; Percomorpha; Tetraodontiformes;
               Tetraodontidae; Tetraodon.
1 (bases 1 to 651)
Roest-Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bertot, A., Fizes, C., Wincker, P., Brottier, P., Quettier, F.,
Saurin, W. and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
REFERENCE       2 (bases 1 to 651)
               Roest-Crolius, H., Jaillon, O., Dasilva, C., Fizes, C., Fisher, C.,
               Bouneau, L., Billault, A., Quettier, F., Saurin, W., Bertot, A. and
               Weissenbach, J.
               Characterization and repeat analysis of the compact genome of the
               freshwater pufferfish Tetraodon nigroviridis
               Unpublished
TITLE
JOURNAL
AUTHORS

```



```

BASE COUNT      136 a   103 c   99 g   116 t
ORIGIN
Query Match      23.9%; Score 37; DB 10; Length 454;
Best Local Similarity 62.4%; Pred. No. 5.2;
Matches 58; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 40 CATGTAGAAGTTGCTGCACCTCCAGCACACCCAGCCGCCAGAGAAAGTGGGGCTCCT 99
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 306 CTTCCAGGAGCTGCTGGACAGCTGTCACAAACCAGGAAACCCAGGAAACCAACCA 365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 100 GGTGCACACAGGTTTACCACAATATACAGAGAA 132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 366 GGTGCACACAGGACTTCCAGGAAACCCAGGAAA 398
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
AW675997
LOCUS
DEFINITION
  SWYD25CAU13H04SK Brugia malayi young adult day 25 cdna
  (SAW99MLW-BmyD25) Brugia malayi cDNA clone SWYD25CAU13H04 5', mRNA
sequence.
ACCESSION
  AW675997
VERSION
  AW675997.1 GI:7545483
KEYWORDS
  EST.
SOURCE
  Brugia malayi.
  Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidae;
  Onchocercidae; Brugia.
  1 (bases 1 to 525)
REFERENCE
  1 Williams, S.A.
  Genes expressed in young adult day 25 of Brugia malayi
  Unpublished (1999)
JOURNAL
  MOLECULAR PARASITOLOGY
CONTACT: Steven A. Williams
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pbluescript SK.
Location/Qualifiers
  1..525
  /organism="Brugia malayi"
  /db_xref="taxon:6279"
  /clone="SWYD25CAU13H04"
  /clone_lib="Brugia malayi young adult day 25 cdna
  (SAW99MLW-BmyD25)"
  /dev_stage="young adult, twenty five days after infection"
  /lab_host="XLI-Blue MRF"
  /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
  Xho I; Lymphatic filarial nematode parasite of humans.
  mRNA was prepared from young adult worms isolated from
  the peritoneal cavity of jirds on day 25 after infection
  and converted to double-stranded cDNA using reverse
  transcriptase and oligo(dT) followed by RNase H and DNA
  pol I. The library has 6.2 x 105 independent recombinants
  and the average insert size is approx.110bp. The library
  was constructed by Michelle Lizotte-Waniewski. The
  library is available from Dr. S.A. Williams, email:
  genome@neal.smith.edu."

BASE COUNT      156 a   118 c   127 g   123 t   1 others
ORIGIN
Query Match      23.7%; Score 36.8; DB 10; Length 525;
Best Local Similarity 61.5%; Pred. No. 5.9;
Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 45 AGAAGTTCTGACCTCCAGCACACCCAGCCGCCAGAGAGTGGGCCCTCTCTGGTGC 104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 329 AGGACCTGATGGCCCCCAGGAAAAACCCAGGGTTCACGAGGACGACGACGAGCGCC 388
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
AW179710
LOCUS
DEFINITION
  SWYD25CAU02H05SK Brugia malayi young adult day 25 cdna
  (SAW99MLW-BmyD25) Brugia malayi cDNA clone SWYD25CAU02H05 5', mRNA
sequence.
ACCESSION
  AW179710
VERSION
  AW179710.1 GI:6445747
KEYWORDS
  EST.
SOURCE
  Brugia malayi.
  Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidae;
  Onchocercidae; Brugia.
  1 (bases 1 to 575)
REFERENCE
  1 Williams, S.A.
  Genes expressed in young adult day 25 of Brugia malayi
  Unpublished (1999)
JOURNAL
  MOLECULAR PARASITOLOGY
CONTACT: Steven A. Williams
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pbluescript SK.
Location/Qualifiers
  1..575
  /organism="Brugia malayi"
  /db_xref="taxon:6279"
  /clone="SWYD25CAU02H05"
  /clone_lib="Brugia malayi young adult day 25 cdna
  (SAW99MLW-BmyD25)"
  /dev_stage="young adult, twenty five days after infection"
  /lab_host="XLI-Blue MRF"
  /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
  Xho I; Lymphatic filarial nematode parasite of humans.
  mRNA was prepared from young adult worms isolated from
  the peritoneal cavity of jirds on day 25 after infection
  and converted to double-stranded cDNA using reverse
  transcriptase and oligo(dT) followed by RNase H and DNA
  pol I. The library has 6.2 x 105 independent recombinants
  and the average insert size is approx.110bp. The library
  was constructed by Michelle Lizotte-Waniewski. The
  library is available from Dr. S.A. Williams, email:
  genome@neal.smith.edu."

BASE COUNT      163 a   144 c   142 g   123 t   3 others
ORIGIN
Query Match      23.7%; Score 36.8; DB 10; Length 575;
Best Local Similarity 61.5%; Pred. No. 5.9;
Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 45 AGAAGTTCTGACCTCCAGCACACCCAGCCGCCAGAGAGTGGGCCCTCTCTGGTGC 104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 311 AGGACCTGATGGCCCCCAGGAAAAACCCAGGGTTCACGAGGACGACGACGAGCGCC 370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 105 ACCAGGTTTACCACAATATATACAGAGAAATAGTGA 140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 371 ACCCGGTTTTCCAGGAGACCAACCAAGAATATGTGA 406
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
AW179840
LOCUS
DEFINITION
  SWYD25CAU04D10SK Brugia malayi young adult day 25 cdna
  (SAW99MLW-BmyD25) Brugia malayi cDNA clone SWYD25CAU04D10 5', mRNA
sequence.
ACCESSION
  AW179840
VERSION
  AW179840.1 GI:6445747
KEYWORDS
  EST.
SOURCE
  Brugia malayi.
  Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidae;
  Onchocercidae; Brugia.
  1 (bases 1 to 575)
REFERENCE
  1 Williams, S.A.
  Genes expressed in young adult day 25 of Brugia malayi
  Unpublished (1999)
JOURNAL
  MOLECULAR PARASITOLOGY
CONTACT: Steven A. Williams
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pbluescript SK.
Location/Qualifiers
  1..575
  /organism="Brugia malayi"
  /db_xref="taxon:6279"
  /clone="SWYD25CAU04D10"
  /clone_lib="Brugia malayi young adult day 25 cdna
  (SAW99MLW-BmyD25)"
  /dev_stage="young adult, twenty five days after infection"
  /lab_host="XLI-Blue MRF"
  /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
  Xho I; Lymphatic filarial nematode parasite of humans.
  mRNA was prepared from young adult worms isolated from
  the peritoneal cavity of jirds on day 25 after infection
  and converted to double-stranded cDNA using reverse
  transcriptase and oligo(dT) followed by RNase H and DNA
  pol I. The library has 6.2 x 105 independent recombinants
  and the average insert size is approx.110bp. The library
  was constructed by Michelle Lizotte-Waniewski. The
  library is available from Dr. S.A. Williams, email:
  genome@neal.smith.edu."

BASE COUNT      156 a   118 c   127 g   123 t   1 others
ORIGIN
Query Match      23.7%; Score 36.8; DB 10; Length 525;
Best Local Similarity 61.5%; Pred. No. 5.9;
Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 45 AGAAGTTCTGACCTCCAGCACACCCAGCCGCCAGAGAGTGGGCCCTCTCTGGTGC 104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 329 AGGACCTGATGGCCCCCAGGAAAAACCCAGGGTTCACGAGGACGACGACGAGCGCC 388
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

sequence.
AW179840
VERSION AW179840.1 GI:6445877
KEYWORDS EST.
SOURCE Brugia malayi.
ORGANISM Brugia malayi.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Brugia.
1 (bases 1 to 585)
Williams,S.A.
Genes expressed in young adult day 25 of Brugia malayi
Unpublished (1999)
Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK.
Location/Qualifiers
1..585
/organism="Brugia malayi"
/db_xref="taxon:6279"
/clone="SWYD25CAU04D10"
/clone_lib="Brugia malayi young adult day 25 cDNA
(SAW99MLW-BmYD25)"
/dev_stage="young adult, twenty five days after infection"
/lab_host="XL1-Blue MRF'"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Lymphatic filarial nematode parasite of humans.
mRNA was prepared from young adult worms isolated from
the peritoneal cavity of birds on day 25 after infection
and converted to double-stranded cDNA using reverse
transcriptase and oligo(dT) followed by RNase H and DNA
pol I. The library has 6.2 x 105 independent recombinants
and the average insert size is approx 1101bp. The library
was constructed by Michelle Lizotte-Waniewski. The
library is available from Dr. S.A. Williams, email:
genome@neal.smith.edu."
170 a 141 c 145 g 128 t 1 others
BASE COUNT
ORIGIN

```

[illegible]

Search completed: February 20, 2003, 06:14:25  
Job time : 395.859 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 18:01:21 ; Search time 12.4727 Seconds  
(without alignments)  
6329.459 Million cell updates/sec

Title: US-09-997-610-1\_COPY\_2\_156

Perfect score: 155

Sequence: 1 atagtgtctactctctt.....agtgaaatgacaaatgcc 155

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 424239 seqs, 254661826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	155	100.0	1381	9	US-09-997-610-1
2	155	100.0	1731	9	US-09-997-610-5
3	111.2	71.7	1377	9	US-09-997-610-3
4	111.2	71.7	1731	9	US-09-997-610-7
5	35.8	23.1	3477	9	US-09-935-868-25
6	35.8	23.1	3507	9	US-09-935-868-23
7	35.2	22.7	1234	9	US-09-954-531-1366
8	35	22.6	3226	10	US-09-954-456-725
9	33.4	21.5	392	10	US-09-960-352-1786
10	33.4	21.5	415	10	US-09-960-352-9137
11	33.2	21.4	549	10	US-09-923-779-49
12	33.2	21.4	561	10	US-09-923-779-47
13	33.2	21.4	617	10	US-09-923-779-51
14	33.2	21.4	648	10	US-09-923-779-66
15	33.2	21.4	653	10	US-09-923-779-53
16	33.2	21.4	655	10	US-09-923-779-68
17	33.2	21.4	656	10	US-09-923-779-69
18	33.2	21.4	658	10	US-09-923-779-71
19	33.2	21.4	659	10	US-09-923-779-52

c	20	33.2	21.4	780	10	US-09-923-779-5	Sequence 5, Appli
	21	33.2	21.4	823	10	US-09-923-779-28	Sequence 28, Appli
	22	33.2	21.4	886	10	US-09-923-779-4	Sequence 4, Appli
	23	32.8	21.2	6158	10	US-09-919-497-6	Sequence 6, Appli
	24	32.8	21.2	6158	10	US-09-954-456-762	Sequence 762, App
	25	32	20.6	990	9	US-09-738-626-2514	Sequence 2514, Ap
c	26	32	20.6	3309400	9	US-09-738-626-1	Sequence 1, Appli
	27	31.6	20.4	1344	10	US-09-925-299-44	Sequence 44, Appli
	28	31.4	20.3	9388	9	US-09-924-400-141	Sequence 141, App
	29	31.4	20.3	9388	10	US-09-810-936-141	Sequence 141, App
	30	31.4	20.3	9388	10	US-09-429-755-141	Sequence 141, App
	31	31.2	20.1	1619	9	US-09-764-868-400	Sequence 400, App
c	32	30.4	19.6	409	12	US-10-044-090-524	Sequence 524, App
	33	30.4	19.6	821	9	US-09-729-658B-7	Sequence 7, Appli
	34	30.4	19.6	1176	9	US-09-729-658B-14	Sequence 14, Appli
	35	30.4	19.6	1574	9	US-09-729-658B-1	Sequence 1, Appli
	36	30.4	19.6	2685	9	US-09-764-868-48	Sequence 48, Appli
	37	30.4	19.6	3380	10	US-09-799-799-1	Sequence 1, Appli
	38	30.4	19.6	3394	10	US-09-880-107-2178	Sequence 2178, Ap
	39	30.2	19.5	162	10	US-09-783-590-11563	Sequence 11563, A
	40	30.2	19.5	88191	10	US-09-799-799-3	Sequence 3, Appli
	41	29.6	19.1	2823	10	US-09-919-497-7	Sequence 7, Appli
	42	29.6	19.1	32183	10	US-09-764-869-1494	Sequence 1494, Ap
c	43	29.4	19.0	1320	10	US-09-815-242-7615	Sequence 7615, Ap
	44	29.2	18.8	2379	9	US-10-063-547-123	Sequence 123, App
	45	29.2	18.8	2379	9	US-10-038-072-535	Sequence 535, App

## ALIGNMENTS

## RESULT 1

US-09-997-610-1  
; Sequence 1, Application US/09997610  
; Patent No. US20020156244A1  
; GENERAL INFORMATION:  
; APPLICANT: Fox, Brian  
; APPLICANT: Holloway, James L.  
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN  
; TITLE OF INVENTION: ZACRP13  
; FILE REFERENCE: 00-96  
; CURRENT APPLICATION NUMBER: US/09/997,610  
; CURRENT FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: US 60/253,924  
; PRIOR FILING DATE: 2000-11-29  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 1381  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (2)...(1381)  
US-09-997-610-1

Query Match 100.0%; Score 155; DB 9; Length 1381;  
Best Local Similarity 100.0%; Pred. No. 3.4e-41;  
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ATAGTGGTTCATACCTGCTTCTTAATACGGCAGTCATTTGACGATGTAGAAAGTTGCTGGACCT	60
Db	2	ATAGTGGTTCATACCTGCTTCTTAATACGGCAGTCATTTGACGATGTAGAAAGTTGCTGGACCT	61
Qy	61	CCAGCACACCCCGCCGCCAGAGAGAGTGGGGCTCTCTGTGACACAGGTTTACCAAA	120
Db	62	CCAGCACACCCCGCCGCCAGAGAGAGTGGGGCTCTCTGTGACACAGGTTTACCAAA	121
Qy	121	TATACAGAGAAATGAAGTGAATGACAAATGCC	155
Db	122	TATACAGAGAAATGAAGTGAATGACAAATGCC	156

## RESULT 2

US-09-997-610-5  
; Sequence 5, Application US/09997610  
; Patent No. US20020156244A1  
; GENERAL INFORMATION:  
; APPLICANT: Fox, Brian  
; APPLICANT: Holloway, James L.  
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN  
; TITLE OF INVENTION: ZACRP13  
; FILE REFERENCE: 00-96  
; CURRENT APPLICATION NUMBER: US/09/997,610  
; CURRENT FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: US 60/253,924  
; PRIOR FILING DATE: 2000-11-29  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 1731  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1731)  
US-09-997-610-5

Query Match 100.0%; Score 155; DB 9; Length 1731;  
Best Local Similarity 100.0%; Pred. No. 3.7e-41;  
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATAGTGGTCATACCTGCTCTTAATAACGGCAGTCATTGAGCATGTAGAGTTGCTGGACCT 60  
|||||  
DB 355 ATAGTGGTCATACCTGCTCTTAATAACGGCAGTCATTGAGCATGTAGAGTTGCTGGACCT 414  
|||||  
QY 61 CCAGCACACCCCGAGCCGCCAGGAAGTGGGGCCTCTCTGGTGCACCAAGTTTACCACAA 120  
|||||  
DB 415 CCAGCACACCCCGAGCCGCCAGGAAGTGGGGCCTCTCTGGTGCACCAAGTTTACCACAA 474  
|||||

QY 121 TATACAGGAGAAATAAGTGAATGACAAAATGCC 155  
|||||  
DB 475 TATACAGGAGAAATAAGTGAATGACAAAATGCC 509  
|||||

## RESULT 3

US-09-997-610-3  
; Sequence 3, Application US/09997610  
; Patent No. US20020156244A1  
; GENERAL INFORMATION:  
; APPLICANT: Fox, Brian  
; APPLICANT: Holloway, James L.  
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN  
; TITLE OF INVENTION: ZACRP13  
; FILE REFERENCE: 00-96  
; CURRENT APPLICATION NUMBER: US/09/997,610  
; CURRENT FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: US 60/253,924  
; PRIOR FILING DATE: 2000-11-29  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 1377  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Degenerate polynucleotide encoding a polypeptide  
; OTHER INFORMATION: of SEQ ID NO:2  
; NAME/KEY: variation  
; LOCATION: (1)...(1377)  
; OTHER INFORMATION: Each n is independently A, T, G, or C.  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1377)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-997-610-3

Query Match 71.7%; Score 111.2; DB 9; Length 1377;  
Best Local Similarity 64.5%; Pred. No. 7.3e-27;  
Matches 100; Conservative 22; Mismatches 33; Indels 0; Gaps 0;  
QY 1 ATAGTGGTCATACCTGCTCTTAATAACGGCAGTCATTGAGCATGTAGAGTTGCTGGACCT 60  
|||||  
DB 1 ATHGTNGTNATHCCNGTNTYNATHACNGCGTNGTNGATGARGTNGCNGNCCN 60  
|||||  
QY 61 CCAGCACACCCCGAGCCGCCAGGAAGTGGGGCCTCTCTGGTGCACCAAGTTTACCACAA 120  
|||||  
DB 61 CCAGCNCAYCCNMGNCCNCGARGGTNGCNGCNCNGCNGCNGGNTTNCNCAR 120  
|||||  
QY 121 TATACAGGAGAAATAAGTGAATGACAAAATGCC 155  
|||||  
DB 121 TAYACNGGNGARATHWSNGARATGACNAARTGYCC 155  
|||||

## RESULT 4

US-09-997-610-7  
; Sequence 7, Application US/09997610  
; Patent No. US20020156244A1  
; GENERAL INFORMATION:  
; APPLICANT: Fox, Brian  
; APPLICANT: Holloway, James L.  
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN  
; TITLE OF INVENTION: ZACRP13  
; FILE REFERENCE: 00-96  
; CURRENT APPLICATION NUMBER: US/09/997,610  
; CURRENT FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: US 60/253,924  
; PRIOR FILING DATE: 2000-11-29  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 7  
; LENGTH: 1731  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Degenerate polynucleotide sequence of zacrp13/zhpl  
; OTHER INFORMATION: of SEQ ID NO:6  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1731)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-997-610-7

Query Match 71.7%; Score 111.2; DB 9; Length 1731;  
Best Local Similarity 64.5%; Pred. No. 8.1e-27;  
Matches 100; Conservative 22; Mismatches 33; Indels 0; Gaps 0;

QY 1 ATAGTGGTCATACCTGCTCTTAATAACGGCAGTCATTGAGCATGTAGAGTTGCTGGACCT 60  
|||||  
DB 355 ATHGTNGTNATHCCNGTNTYNATHACNGCGTNGTNGATGARGTNGCNGNCCN 414  
|||||  
QY 61 CCAGCACACCCCGAGCCGCCAGGAAGTGGGGCCTCTCTGGTGCACCAAGTTTACCACAA 120  
|||||  
DB 415 CCAGCNCAYCCNMGNCCNCGARGGTNGCNGCNCNGCNGCNGGNTTNCNCAR 474  
|||||  
QY 121 TATACAGGAGAAATAAGTGAATGACAAAATGCC 155  
|||||  
DB 475 TAYACNGGNGARATHWSNGARATGACNAARTGYCC 509  
|||||

## RESULT 5

US-09-935-868-25  
; Sequence 25, Application US/09935868  
; Patent No. US20020164690A1  
; GENERAL INFORMATION:  
; APPLICANT: Regeneron Pharmaceuticals, Inc  
; TITLE OF INVENTION: Receptor Based Antagonists, and Methods of Making and Using  
; FILE REFERENCE: REG 203D  
; CURRENT APPLICATION NUMBER: US/09/935,868  
; CURRENT FILING DATE: 2002-04-11  
; PRIOR APPLICATION NUMBER: PCT/US99/22045



```
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 725
; LENGTH: 3226
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-725

Query Match      22.6%; Score 35; DB 10; Length 3226;
Best Local Similarity 61.5%; Pred. No. 0.096;
Matches 56; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 26 CGCGAGTCATTGAGCATGTAGAAAGTTGCTGGACCTCCAGCACACCCAGGCCCCAGGAGAG 85
Db 794 CAGAAGGCATTGGAAGCCAGGAGCTGCTGGAGCCCGCCAGGCCAGCGAGGATTCCAGGAA 853

Qy 86 AAGTGGGGCCTCTCGTGGTCACACAGGTTTACC 116
Db 854 CAAAAGGTCCTCCCTGGGGCTCCAGGAATAGC 884

RESULT 9
US-09-960-352-1786
; Sequence 1786, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 1786
; LENGTH: 392
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 08-LIB34-023-Q1-E1-B7
US-09-960-352-1786

Query Match      21.5%; Score 33.4; DB 10; Length 392;
Best Local Similarity 62.7%; Pred. No. 0.12;
Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 46 GAAGTTGCTGGACCTCCAGCACACCCAGGCCCCAGCAAGAAGTGGGGCCTCCTCGTGTCGA 105
Db 146 GGACCTCGGGGCCACAGGACACCCAGGGCCTCCAGGTATAGAAGAGGGCCTCCAGGTATA 205

Qy 106 CCAGGTTTACCACATATACAGG 128
Db 206 AGAGGAATACCAGGTTTGCAGG 228

RESULT 10
US-09-960-352-9137
; Sequence 9137, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 9137
; LENGTH: 415
```

```
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 39-LIB34-018-Q1-E1-B12
US-09-960-352-9137

Query Match      21.5%; Score 33.4; DB 10; Length 415;
Best Local Similarity 62.7%; Pred. No. 0.13;
Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 46 GAAGTTGCTGGACCTCCAGCACACCCAGGCCCCAGCAAGAAGTGGGGCCTCCTCGTGTCGA 105
Db 156 GGACCTCGGGGCCACAGGACACCCAGGGCCTCCAGGTATAGAAGAGGGCCTCCAGGTATA 215

Qy 106 CCAGGTTTACCACATATACAGG 128
Db 216 AGAGGAATACCAGGTTTGCAGG 238

RESULT 11
US-09-923-779-49
; Sequence 49, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 480, 498, 523, 539
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-49

Query Match      21.4%; Score 33.2; DB 10; Length 549;
Best Local Similarity 54.0%; Pred. No. 0.17;
Matches 68; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 16 GTCTTAATAACGGCAGTCATTGAGCATGTAGAAAGTTGCTGGACCTCCAGCACACCCAGG 75
Db 313 GTAGCAGTAGGAGGACTCGCAGGCTATCTCTGGACCAGCTGGCCCCCAGGCCCTCCCGGT 372

Qy 76 CCCCCAAGAAGTGGGGCCTCCTCGTGCACCCAGGTTTACCACATATATACAGGAGAAATA 135
Db 373 CCCCTGGTACATCTGGTCATCTCTGTTCCCTCGTATCTCCAGGTATCCAGGATACCAAGGACCCCT 432

Qy 136 AGTGAA 141
Db 433 GGTGAA 438

RESULT 12
US-09-923-779-47
; Sequence 47, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
```

```
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 544, 550
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-47

Query Match      21.4%; Score 33.2; DB 10; Length 561;
Best Local Similarity 61.6%; Pred. No. 0.17;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 53 CTGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGCCCTCCCTGGTGACACAGGTT 112
      ||||| ||||| ||| | ||||| | ||| ||||| ||||| ||||| |||||
Db 404 CTGGATCTCCAGGATACCAAGGACCCCTCGTGAACCTGGCAAGCTGGTCTTCAGGCC 463
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 113 TACCACAATATACAGGAGAAATAAGT 138
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 464 CTCGAGGACCTCCTGGTGCTATAGT 489

RESULT 13
US-09-923-779-51
; Sequence 51, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51
; LENGTH: 617
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 581, 605
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-51

Query Match      21.4%; Score 33.2; DB 10; Length 617;
Best Local Similarity 61.6%; Pred. No. 0.18;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 53 CTGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGCCCTCCCTGGTGACACAGGTT 112
      ||||| ||||| ||| | ||||| | ||| ||||| ||||| ||||| |||||
Db 404 CTGGATCTCCAGGATACCAAGGACCCCTCGTGAACCTGGCAAGCTGGTCTTCAGGCC 463
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 113 TACCACAATATACAGGAGAAATAAGT 138
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 464 CTCGAGGACCTCCTGGTGCTATAGT 489

RESULT 14
US-09-923-779-66
; Sequence 66, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
```

```
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 648
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 642, 646
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-66

Query Match      21.4%; Score 33.2; DB 10; Length 648;
Best Local Similarity 61.6%; Pred. No. 0.18;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 53 CTGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGCCCTCCCTGGTGACACAGGTT 112
      ||||| ||||| ||| | ||||| | ||| ||||| ||||| ||||| |||||
Db 404 CTGGATCTCCAGGATACCAAGGACCCCTCGTGAACCTGGCAAGCTGGTCTTCAGGCC 463
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 113 TACCACAATATACAGGAGAAATAAGT 138
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 464 CTCGAGGACCTCCTGGTGCTATAGT 489

RESULT 15
US-09-923-779-53
; Sequence 53, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 53
; LENGTH: 653
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 517, 579, 581, 603, 649
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-53

Query Match      21.4%; Score 33.2; DB 10; Length 653;
Best Local Similarity 61.6%; Pred. No. 0.18;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 53 CTGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGCCCTCCCTGGTGACACAGGTT 112
      ||||| ||||| ||| | ||||| | ||| ||||| ||||| ||||| |||||
Db 404 CTGGATCTCCAGGATACCAAGGACCCCTCGTGAACCTGGCAAGCTGGTCTTCAGGCC 463
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 113 TACCACAATATACAGGAGAAATAAGT 138
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 464 CTCGAGGACCTCCTGGTGCTATAGT 489

Search completed: February 19, 2003, 23:02:39
Job time : 20.4727 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 16:09:30 ; Search time 10.2324 Seconds  
(without alignments)  
4645.518 Million cell updates/sec

Title: us-09-997-610-1\_COPY\_2\_156

Perfect score: 155  
Sequence: 1 atagtgttcatacctgtctt.....agtgaaatgacaaatgcc 155

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35.8	23.1	1074	2	US-08-627-151A-15
2	35.8	23.1	1404	6	Sequence 15, Appli
3	35.8	23.1	1404	6	Patent No. 5171840
4	35.8	23.1	1486	4	US-08-795-473B-3
5	35.8	23.1	1486	4	Sequence 3, Appli
6	35.8	23.1	2061	6	US-09-439-856-3
7	35.8	23.1	2061	6	Patent No. 5171840
8	35.8	23.1	3319	4	US-08-795-473B-2
9	35.8	23.1	3319	4	Sequence 2, Appli
10	35.8	23.1	1881	4	US-09-439-856-2
11	31.4	20.3	1395	4	Sequence 20, Appli
12	31.4	20.3	9388	4	US-09-029-348-20
13	31.4	20.3	9388	4	Sequence 88, Appli
14	31.4	20.3	9388	4	Sequence 141, App
15	30.8	19.9	3294	4	US-08-991-789A-141
16	30.4	19.6	821	4	US-09-052-451-141
17	30.4	19.6	1176	4	US-09-598-326-141
18	30.4	19.6	1574	4	US-08-923-992A-7
19	30.4	19.6	3394	1	US-09-342-681C-7
20	29.4	19.0	2543	1	US-09-342-681C-14
21	29.4	19.0	2543	3	US-08-159-784-4
22	29.2	18.8	3492	4	US-08-555-669-11
23	29.2	18.8	4200	1	US-09-073-663-11
24	29.2	18.8	4200	1	US-08-923-992A-9
25	29.2	18.8	4200	1	US-08-242-932-1
26	29.2	18.8	4200	5	US-08-714-481-1
27	28.8	18.6	1898	1	Sequence 1, Appli
28	28.8	18.6	1898	1	Sequence 1, Appli
29	28.8	18.6	1898	1	Sequence 1, Appli
30	28.8	18.6	1898	1	Sequence 1, Appli
31	28.8	18.6	1898	1	Sequence 1, Appli
32	28.8	18.6	1898	1	Sequence 1, Appli
33	28.8	18.6	1898	1	Sequence 1, Appli
34	28.8	18.6	1898	1	Sequence 1, Appli
35	28.8	18.6	1898	1	Sequence 1, Appli
36	28.8	18.6	1898	1	Sequence 1, Appli
37	28.8	18.6	1898	1	Sequence 1, Appli
38	28.8	18.6	1898	1	Sequence 1, Appli
39	28.8	18.6	1898	1	Sequence 1, Appli
40	28.8	18.6	1898	1	Sequence 1, Appli
41	28.8	18.6	1898	1	Sequence 1, Appli
42	28.8	18.6	1898	1	Sequence 1, Appli
43	28.8	18.6	1898	1	Sequence 1, Appli
44	28.8	18.6	1898	1	Sequence 1, Appli
45	28.8	18.6	1898	1	Sequence 1, Appli

Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 635, App  
Sequence 647, App  
Sequence 726, App  
Sequence 624, App  
Sequence 16, Appl  
Sequence 693, App  
Sequence 693, App  
Sequence 139, App  
Sequence 5, Appli  
Sequence 5, Appli  
Patent No. 5510466  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 1, Appli  
Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-08-627-151A-15  
; Sequence 15, Application US/08627151A  
; Patent No. 5866341  
; GENERAL INFORMATION:  
; APPLICANT: SPINELLA, Dominic  
; APPLICANT: BECHERER, Kathleen  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; SCREENING DRUG LIBRARIES  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Gen-Probe Incorporated  
; STREET: 10210 Genetic Center Drive  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92121  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/627,151A  
; FILING DATE: 03-APR-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fisher, Carlos A  
; REGISTRATION NUMBER: 36,510  
; REFERENCE/DOCKET NUMBER: CBI016  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-410-8926  
; TELEFAX: 619-410-8928  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1074 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-627-151A-15

Query Match 23.1% Score 35.8; DB 2; Length 1074;  
Best Local Similarity 54.1%; Pred. NO. 0.026; Indels 0; Gaps 0;  
Matches 73; Conservative 0; Mismatches 62;





```
; NAME: Davidson, Clifford M.
; REGISTRATION NUMBER: 32,728
; REFERENCE/DOCKET NUMBER: 963.1007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)-997-1028
; TELEFAX: (212)-997-1037
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3319 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
US-08-795-473B-2

Query Match      23.1%; Score 35.8; DB 4; Length 3319;
Best Local Similarity 54.1%; Pred. No. 0.04;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 10 ATACCTGTCTTAATAACGGCAGTCATTGAGCATGTAGAGTTGCTGGACCTCCAGCACAC 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1205 ATATCGGGCTGAACGGTCAAGACATTCACAACATGGATGGTCAAGGACCTCCAGCATCA 1264

QY 70 CCCAGGCCCCAGAGAAGTGGGGCTCTGTGTGCACACAGGTTTACCACAATATACAGGA 129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1265 CTGTGTCAATCCACGACGCTGGAGGGCTGTAGGCACGTGGTGCAGCTTCGTGCCCAGGA 1324

QY 130 GAAATAAGTGAATG 144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1325 GGAGTTCGGGCAAGG 1339

RESULT 9
US-09-439-856-2
; Sequence 2, Application US/09439856
; Patent No. 641009
; GENERAL INFORMATION:
; APPLICANT: Galun, Eithan
; APPLICANT: Nahot, Orit
; APPLICANT: Blum, Herbert E.
; TITLE OF INVENTION: A Pharmaceutical Composition for Treating
; TITLE OF INVENTION: Hepatitis B Virus (HBV) Infection
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; STREET: Davidson, Davidson and Kappel, LLC
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS-DOS EDITOR
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/439,856
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/795,473
; FILING DATE: 11-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Davidson, Clifford M.
; REGISTRATION NUMBER: 32,728
; REFERENCE/DOCKET NUMBER: 963.1007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)-997-1028
; TELEFAX: (212)-997-1037
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3319 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

; NAME: Davidson, Clifford M.
; REGISTRATION NUMBER: 32,728
; REFERENCE/DOCKET NUMBER: 963.1007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)-997-1028
; TELEFAX: (212)-997-1037
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3319 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

; TOPOLOGY: unknown
US-09-439-856-2

Query Match      23.1%; Score 35.8; DB 4; Length 3319;
Best Local Similarity 54.1%; Pred. No. 0.04;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 10 ATACCTGTCTTAATAACGGCAGTCATTGAGCATGTAGAGTTGCTGGACCTCCAGCACAC 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1205 ATATCGGGCTGAACGGTCAAGACATTCACAACATGGATGGTCAAGGACCTCCAGCATCA 1264

QY 70 CCCAGGCCCCAGAGAAGTGGGGCTCTGTGTGCACACAGGTTTACCACAATATACAGGA 129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1265 CTGTGTCAATCCACGACGCTGGAGGGCTGTAGGCACGTGGTGCAGCTTCGTGCCCAGGA 1324

QY 130 GAAATAAGTGAATG 144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1325 GGAGTTCGGGCAAGG 1339

RESULT 10
US-09-029-348-20
; Sequence 20, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
; FILE REFERENCE: D087857/PUS LISTING
; CURRENT APPLICATION NUMBER: US/09/029,348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 1881
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
; OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS
US-09-029-348-20

Query Match      21.4%; Score 33.2; DB 4; Length 1881;
Best Local Similarity 54.0%; Pred. No. 0.22;
Matches 68; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 16 GTCTTAATAACGGCAGTCATTGAGCATGTAGAGTTGCTGGACCTCCAGCACACCCACGG 75
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 490 GTAGCAGTAGGAGGACTCGCAGGCTATCTGGACCAGCTGGCCGCCCGCCGCTCCCGGT 549

QY 76 CCCCCAGAGAAGTGGGGCTCTGTGTGCACACAGGTTTACCACAATATACAGGAGAAATA 135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 550 CCCCCGTGTACATCTGGTCTCTCTGTTCCCTGGATCTCCAGGATACCAAGAGCCCCCT 609

QY 136 AGTGAA 141
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 610 GGTGAA 615

RESULT 11
US-09-149-476-88/c
; Sequence 88, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002PI
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
```



EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,614  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,578  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,576  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/047,501  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,670  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/056,632  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,664  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,876  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,881  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,909  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,875  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,862  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,887  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,908  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/048,964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/057,650  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/056,884  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057,669  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/049,610  
EARLIER FILING DATE: 1997-06-13  
EARLIER APPLICATION NUMBER: 60/061,060  
EARLIER FILING DATE: 1997-10-02

Query Match 20.3%; Score 31.4; DB 4; Length 1395;  
Best Local Similarity 51.8%; Pred. No. 0.74;  
Matches 71; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 16 GCTTAATAACGGCAGTCATTGAGCATGTAGAACTTGTGGACCTCCACACACCCAGG 75  
DB 1311 GGCTAACGCTCGAAATTCCTCGGCTGAAGAGGGGCTAGATTGCTTTTATCTTTGG 1252  
QY 76 CCCCCAGAGAAGTGGGGCTCTGCTGCACCCAGGTTTACCAACAATATACAGGAGAAATA 135  
DB 1251 TTAGAAGGAGAGGGGGTCTAGTTAAACAATTTTACAGAAATAAAGTAGGCAAAA 1192  
QY 136 AGTGAATGACAAATG 152  
DB 1191 AGTAAAGGATAAATG 1175

RESULT 12  
US-08-991-789A-141/c  
; Sequence 141, Application US/08991789A  
; Patent No. 6225054  
; GENERAL INFORMATION:  
; APPLICANT: Frudakis, Tony N.  
; Smith, John M.  
; Reed, Steven G.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TREATMENT AND DIAGNOSIS OF BREAST CANCER  
; NUMBER OF SEQUENCES: 292  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed IP Law Group  
; STREET: 701 Fifth Avenue, Suite 6300

CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/991,789A  
FILING DATE: 11-Dec-1997  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Potter, Jane E. R.  
REGISTRATION NUMBER: 33,332  
REFERENCE/DOCKET NUMBER: 210121.419C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 141:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9388 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 141:  
US-08-991-789A-141

Query Match 20.3%; Score 31.4; DB 4; Length 9388;  
Best Local Similarity 51.8%; Pred. No. 1.5;  
Matches 71; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 16 GCTTAATAACGGCAGTCATTGAGCATGTAGAACTTGTGGACCTCCACACACCCAGG 75  
DB 8230 GGCTAACGCTCGAAATTCCTCGGCTGAAGAGGGGCTAGATTGCTTTTATCTTTGG 8171  
QY 76 CCCCCAGAGAAGTGGGGCTCTGCTGCACCCAGGTTTACCAACAATATACAGGAGAAATA 135  
DB 8170 TTAGAAGGAGAGGGGGTCTAGTTAAACAATTTTACAGAAATAAAGTAGGCAAAA 8111  
QY 136 AGTGAATGACAAATG 152  
DB 8110 AGTAAAGGATAAATG 8094

RESULT 13  
US-09-062-451-141/c  
; Sequence 141, Application US/09062451  
; Patent No. 6344550  
; GENERAL INFORMATION:  
; APPLICANT: Frudakis, Tony N.  
; Smith, John M.  
; Reed, Steven G.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TREATMENT AND DIAGNOSIS OF BREAST CANCER  
; NUMBER OF SEQUENCES: 297  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/062,451  
FILING DATE: 04-APR-1997  
CLASSIFICATION:

```

US-09-598-326-141
Query Match          20.3%; Score 31.4; DB 4; Length 9388;
Best Local Similarity 51.8%; Pred. No. 1.5;
Matches 71; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 16 GTCCTAATAACGGCAGTCATTTGAGCATGTGAGAGTTGCTCGGACCTCCAGCACACACCCAGG 75
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8230 GGCTAAGCGCTCGAAATTTCTCTCGGCCCTGAAGAAGGGGCTAGATTTTGCTTTTATACTTTGG 8171

Qy 76 CCCCCAGAAGAAGTGGGGCCCTCCTGTGTGCACCGAGTTTACCACAAATATACAGGAGAATA 135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8170 TTTAGAAGGAGAGGGGGGTCTAGTTAAAACAATTTTACAGAAATAAAGTAGGCAAAAA 8111

Qy 136 AGTGAATGACAAAATG 152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8110 AGTTAAAAGGATAATG 8094

RESULT 15
US-08-923-992A-7
; Sequence 7, Application US/08923992A
; Patent No. 6280738
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-IGA Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,992A
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3294 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3294
; US-08-923-992A-7

Query Match          19.9%; Score 30.8; DB 4; Length 3294;
Best Local Similarity 57.1%; Pred. No. 1.6;
Matches 56; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Qy 22 ATAACGGCAGTCATTTGAGCATGTGAGAGTTGCTCGGACCTCCAGCACACACCCAGCCCCA 81
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 2626 ATTCCGGAAACCCCTAAGACTCCAGAGCTCCCTAACCTTCCAGACGTCCCTTAAGCTTCA 2685

Qy 82 GAAGAAGTGGGGCTCCTCGTGGTGCACCAAGGTTTACCACA 119

Db 2686 GACGTCCTAAGCTTCCAGATGCACCGAAGTTACCACA 2723

Search completed: February 19, 2003, 22:59:00  
Job time : 21.2324 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 15:08:07 ; Search time 53.6445 Seconds  
(without alignments)  
6506.903 Million cell updates/sec

Title: US-09-997-610-1\_COPY\_2\_156  
Perfect score: 155  
Sequence: 1 atagtgctactactgtctt.....agtgaatgacaaaatgcc 155

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues 4370478  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :				N.Geneseq_101002.*			
1:	/SID52/gcgdata/genseq/genseq-emb1/NA1980.DAT.*	2:	/SID52/gcgdata/genseq/genseq-emb1/NA1981.DAT.*	3:	/SID52/gcgdata/genseq/genseq-emb1/NA1982.DAT.*	4:	/SID52/gcgdata/genseq/genseq-emb1/NA1983.DAT.*
5:	/SID52/gcgdata/genseq/genseq-emb1/NA1984.DAT.*	6:	/SID52/gcgdata/genseq/genseq-emb1/NA1985.DAT.*	7:	/SID52/gcgdata/genseq/genseq-emb1/NA1986.DAT.*	8:	/SID52/gcgdata/genseq/genseq-emb1/NA1987.DAT.*
9:	/SID52/gcgdata/genseq/genseq-emb1/NA1988.DAT.*	10:	/SID52/gcgdata/genseq/genseq-emb1/NA1989.DAT.*	11:	/SID52/gcgdata/genseq/genseq-emb1/NA1990.DAT.*	12:	/SID52/gcgdata/genseq/genseq-emb1/NA1991.DAT.*
13:	/SID52/gcgdata/genseq/genseq-emb1/NA1992.DAT.*	14:	/SID52/gcgdata/genseq/genseq-emb1/NA1993.DAT.*	15:	/SID52/gcgdata/genseq/genseq-emb1/NA1994.DAT.*	16:	/SID52/gcgdata/genseq/genseq-emb1/NA1995.DAT.*
17:	/SID52/gcgdata/genseq/genseq-emb1/NA1996.DAT.*	18:	/SID52/gcgdata/genseq/genseq-emb1/NA1997.DAT.*	19:	/SID52/gcgdata/genseq/genseq-emb1/NA1998.DAT.*	20:	/SID52/gcgdata/genseq/genseq-emb1/NA1999.DAT.*
21:	/SID52/gcgdata/genseq/genseq-emb1/NA2000.DAT.*	22:	/SID52/gcgdata/genseq/genseq-emb1/NA2001A.DAT.*	23:	/SID52/gcgdata/genseq/genseq-emb1/NA2001B.DAT.*	24:	/SID52/gcgdata/genseq/genseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59	38.1	1338	24	Human genset metab
2	38	24.5	328	21	Human secreted pro
3	36.8	23.7	6512	24	Mouse ischaemic co
4	35.8	23.1	975	21	Human interleukin
5	35.8	23.1	1035	21	SR345 coding seque
6	35.8	23.1	1074	18	Interleukin 6 rece
7	35.8	23.1	1260	20	Human IL-6 recepto
8	35.8	23.1	1486	19	Human interleukin-
9	35.8	23.1	1545	21	IL-6R/IL-6 fusion

10	35.8	23.1	1627	18	AAT97848	Human fusion polyp
11	35.8	23.1	2061	10	AA90340	Sequence encoding
12	35.8	23.1	2066	14	AAQ41746	IL-6 receptor codi
13	35.8	23.1	2087	17	AA90847	DNA contg. region
14	35.8	23.1	3319	17	AAT31441	Interleukin-6 rece
15	35.8	23.1	3319	19	AAV60295	Human interleukin-
16	35.8	23.1	3319	21	AAF21364	Human low adenosin
17	35.8	23.1	3319	21	AAA35242	Human adenosine re
18	35.8	23.1	3319	24	ABK84527	Human cDNA differe
19	35.8	23.1	3477	21	AAA09047	Fusion polypeptide
20	35.8	23.1	3507	21	AAAO9046	Human low adenosin
21	35.8	23.1	4513	21	AAF21365	Human low adenosin
22	35.8	23.1	4873	21	AAAS5243	Human adenosine re
23	35.2	22.7	1234	24	ABL64096	Breast cancer rela
24	35	22.6	3226	24	ABL64096	Human cDNA differe
25	35	22.6	3226	24	ABL64096	Lung cancer relate
26	34.8	22.5	1121	23	AAAS89670	DNA encoding novel
27	34.8	22.5	5467	22	AAH98343	Human EST-derived
28	34.8	22.5	5468	22	AAH98411	Human EST-derived
29	34.6	22.3	4428	22	AAAD06578	Porcine alpha1(III
30	34.4	22.2	267	24	ABN20477	Human ORFX polynuc
31	34.2	22.1	1518	22	AAAS01474	Human secreted pro
32	34.2	22.1	1730	21	AAAZ6383	Human secreted pro
33	34.2	22.1	2592	20	AAZ41355	Human normal uteru
34	34.2	22.1	3954	21	AACT5733	Human ORFX ORF1288
35	34.2	22.1	4428	22	AAAD06574	Bovine alpha1(III)
36	34.2	22.1	4428	22	AAAD06575	Bovine alpha1(III)
37	34.2	22.1	9287	24	ABK64501	Human benign prost
38	33.4	21.5	433	21	AAC61719	Clone VGP22 of a g
39	33.2	21.4	549	24	ABK44109	cDNA #49 encoding
40	33.2	21.4	561	24	ABK44107	cDNA #47 encoding
41	33.2	21.4	617	24	ABK44111	cDNA #51 encoding
42	33.2	21.4	648	24	ABK44126	cDNA #66 encoding
43	33.2	21.4	653	24	ABK44113	cDNA #53 encoding
44	33.2	21.4	655	24	ABK44128	cDNA #68 encoding
45	33.2	21.4	656	24	ABK44129	cDNA #69 encoding

ALIGNMENTS

RESULT 1	
AAAL4066	
ID	AAAL4066 standard; cDNA; 1338 BP.
XX	
AC	AAAL4066;
XX	
XX	27-SEP-2002 (first entry)
XX	
XX	Human genset metabolic gene (GMG-9) cDNA sequence.
DE	
XX	Human; gene; ss; gene therapy; genset metabolic gene
KW	GMG-8; GMG-9; GMG-10; GMG-11; metabolic-related disorder
KW	impaired glucose tolerance; insulin resistance; Syndrome X
KW	Type II diabetes; hyperlipidaemia; atherosclerosis; hypertension
KW	heart disease; cardiac insufficiency; coronary insufficiency
KW	high blood pressure; insulin sensitiser;
KW	non-insulin dependent diabetes mellitus.
XX	
OS	Homo sapiens.
XX	
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	1..1338
FT	/*tag= a
FT	/partial
FT	/product= "Human GMG-9 protein"
FT	/note= "No stop codon is given"
XX	
XX	WO200255694-A2.
XX	
PD	18-JUL-2002.
XX	
XX	15-JAN-2002; 2002WO-IB01215.

```
XX 16-JAN-2001; 2001US-262235P.
XX (GEST ) GENSET.
XX Erickson MR, Bour BA, Bihain B, Tanaka H;
XX WPI; 2002-557821/59.
XX P-PSDB; AA015423.
XX Treating or preventing a metabolic-related disease or disorder, e.g.
XX obesity, impaired glucose tolerance, insulin resistance, Syndrome X, or
XX Type II diabetes, comprises administering Genset Metabolic Genes -
XX Disclosure; Page 122-124; 128pp; English.
XX The invention comprises the amino acid and coding sequences of six human
XX genset metabolic genes (GMG-7A, GMG-7B, GMG-8, GMG-9, GMG-10 and GMG-11).
XX The GMG DNA and protein sequences of the invention are useful for
XX treating or preventing metabolic-related disorders, such as: obesity;
XX impaired glucose tolerance; insulin resistance; Syndrome X; Type II
XX diabetes; hyperlipidaemia; atherosclerosis; hypertension; and heart
XX diseases (e.g. cardiac insufficiency, coronary insufficiency or high
XX blood pressure). The GMG DNA and protein sequences of the invention may
XX also be used as insulin sensitizers - for improving insulin sensitivity
XX in persons with non-insulin dependent diabetes mellitus. The present cDNA
XX sequence encodes the human GMG-9 protein.
XX Sequence 1338 BP; 359 A; 288 C; 350 G; 341 T; 0 other;
XX
XX Query Match 38.1%; Score 59; DB 24; Length 1338;
XX Best Local Similarity 92.5%; Pred. No. 6e-09;
XX Matches 62; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
XX
QY . 89 TGGGGCCTCTGGTCACAGGTTTACCACAATATACAGGAGAAATAGTGAATGACAA 148
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
50 TGGATGTCTGTGTCGCCAGGTTTACCACAATATACAGGAGAAATAGTGAATGACAA 109
QY 149 AATGCC 155
Db |||||
110 AATGCC 116
XX
RESULT 2
AAC01552
ID AAC01552 standard; cDNA; 328 BP.
XX
XX AAC01552;
AC AAC01552;
XX
XX 06-OCT-2000 (first entry)
DE
DE Human secreted protein 5' EST, SEQ ID NO: 1550.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
OS
XX EP1033401-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 21-FEB-2000; 2000EP-0200610.
PF
XX
XX 26-FEB-1999; 99US-0122487.
PR
XX
XX (GEST ) GENSET.
PA
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI; 2000-500381/45.
DR P-PSDB; AG01546.
XX
XX
```

```
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1; SEQ ID 1550; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. An ORF has been identified within the
XX sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
XX derived from 30 different tissues. EST sequences usually correspond
XX mainly to the 3' untranslated region (UTR) of the mRNA because they are
XX often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
XX well suited for isolating cDNA sequences derived from the 5' ends of
XX mRNAs and even in those cases where longer cDNA sequences have been
XX obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
XX mRNAs with intact 5' ends and can therefore be used to obtain full length
XX cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
XX gene therapy and chromosome mapping procedures. They are used to obtain
XX upstream regulatory sequences and to design expression and secretion
XX vectors.
XX Sequence 328 BP; 76 A; 77 C; 81 G; 88 T; 6 other;
XX
XX Query Match 24.5%; Score 38; DB 21; Length 328;
XX Best Local Similarity 52.2%; Pred. No. 0.02;
XX Matches 71; Conservative 5; Mismatches 60; Indels 0; Gaps 0;
XX
QY 13 CCTGTCTTAAACGCGCATTTGAGCATGTAGAGTTGCTGGACCTCCAGCACACCCC 72
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
61 MCGGTCGCGAMTGTCTCGCGGTTGTGTCAGTGTGAATTTGTACAGCTGCAGTTGCTCC 120
QY 73 AGGCCCCCAAGAAGTGGGGCTCTCTGTGCACCAAGTTTACCACAATATACAGGAGAA 132
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
121 CGCCCCCGAGMARMCGAGAGTCTACCATGGCTCAAGAATCTCCCAAAATTCAGCAGCA 180
QY 133 ATAAGTGAATGACAA 148
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
181 GAAATTCAGTGACTA 196
XX
RESULT 3
ABI99819
ID ABI99819 standard; cDNA; 6512 BP.
XX
XX ABI99819;
AC ABI99819;
XX
XX 07-MAR-2002 (first entry)
DT
XX
XX Mouse ischaemic condition related cDNA sequence SEQ ID NO:932.
DE
XX
XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
KW vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
XX
XX Mus musculus.
OS
XX WO200188188-A2.
PN
XX
XX 22-NOV-2001.
PD
XX
XX 18-MAY-2001; 2001WO-JP04192.
PF
XX
XX 18-MAY-2000; 2000JP-0145977.
PR
XX
XX (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.
PA
XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
XX
XX WPI; 2002-034733/04.
DR P-PSDB; ABB57334.
XX
XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT expression levels of particular genes defined in the specification or
PT by determining the expression profile of a gene group comprising these
```

```
PT genes -
XX Claim 2; Page 2340-2352; 2690pp; English.
XX The present invention describes a method for examining ischaemic
CC conditions, comprising measuring the expression levels of particular
CC genes (I) in a test sample or determining the expression profile of a
CC gene group in the sample comprising genes selected from (I). The method
CC is useful for examining the ischaemic condition (e.g. compressive
CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
CC the expression levels of particular genes (AB199202 to AB199912, encoding
CC the protein sequences in AB199202 to AB199912) or by determining the
CC expression profile of a gene group comprising these genes. The
CC expression levels or expression profiles produced by these genes are
CC used as an indicator when screening for ischaemic condition-improving
CC drugs or therapeutics for ischaemic diseases. AB199913 and AB199914
CC represent PCR primers for a mouse ischaemic condition related sequence,
CC which are used in the exemplification of the present invention.
XX
XX Sequence 6512 BP; 1553 A; 1798 C; 1950 G; 1211 T; 0 other;
SQ
Query Match 23.7%; Score 36.8; DB 24; Length 6512;
Best Local Similarity 61.5%; Pred. No. 0.16;
Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
QY 53 CTGGACCTCCAGCACACCCAGGCCCCAGAGAGAGTGGGGCTCTCTGGTGCCACAGGTT 112
DB 2692 CAGGACAGTCAGGCTCCCTGGCCCTCTGGACAGCAGGAGCACCTGGAGTTCAGGGT 2751
QY 113 TACCACAATATACAGAGAAATAGTGAAATGACAA 148
DB 2752 TCCCAGGTTCTAAGGTGAATGGTGTCTATGGGAA 2787
RESULT 4
AA70701
ID AAA70701 standard; cDNA; 975 BP.
XX
AC AAA70701;
XX
XX 11-DEC-2000 (first entry)
XX
DE Human interleukin 6 receptor cDNA.
XX
XX Human; interleukin-6 receptor; fungus; Pichia pastoris; PCR primer; ss;
KW expression vector; immunoglobulin-like region; cytokine receptor region.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..975
FT FT /*tag= a
FT FT /partial
FT FT /product= "human IL-6R amino acids 20-344"
FT FT /note= "no start or stop codon is given at the 5' or
FT FT 3' ends of the sequence"
FT FT /transl_except= (pos:169..171,aa:Glu)
FT FT /transl_except= (pos:841..843,aa:Gly)
XX
XX JP2000157280-A.
XX
XX 13-JUN-2000.
XX
XX 26-NOV-1998; 98JP-0335464.
XX
XX 26-NOV-1998; 98JP-0335464.
XX
XX (TOYJ ) TOSOH CORP.
XX
XX WPI; 2000-468203/41.
XX
XX P-PSDB; AAB15389.
XX
XX Yeasts transformed with IL-6 receptor gene -
```

```
XX
XX Example 1; Page 6-8; 10pp; Japanese.
XX
XX The invention relates to the production of human interleukin-6 receptor
CC (IL-6R) protein in the fungus Pichia pastoris. The fungus was
CC transformed with an expression vector (pPIC9-A20LL) containing a gene
CC encoding an IL-6R protein having an immunoglobulin-like region and
CC cytokine receptor region and spanning amino acids from Leu30-Ala323. The
CC human IL-6R region was PCR amplified using the primers AAA70702-A70703.
CC This sequence represents the coding region for the IL-6R protein of the
CC invention.
XX
XX Sequence 975 BP; 208 A; 289 C; 290 G; 188 T; 0 other;
SQ
Query Match 23.1%; Score 35.8; DB 21; Length 975;
Best Local Similarity 54.1%; Pred. No. 0.16;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 10 ATACCTGTCTTAATACGGCAGTCATTCAGCATGTAGAGTGTGCGACCTCCAGCACAC 69
DB 711 ATATCGGGCTGAACGGTCAAAAGACATTCAACATGGTCAAGGACCTCCAGCATCA 770
QY 70 CCCAGGCCCCAGAGAAAGTGGGGCTCTCTGTGTGCACACAGGTTTACCACAATATACAGGA 129
DB 771 CTGTGTATCCACACGCGCTCGAGGGGCTGAGGCACGCTGTGCGACGTTCTGTCGCCACGA 830
QY 130 GAAATAAGTGAATG 144
DB 831 GGAGTTCGGGCAAGG 845
RESULT 5
AAZ40288
ID AAZ40288 standard; DNA; 1035 BP.
XX
AC AAZ40288;
XX
XX 25-FEB-2000 (first entry)
XX
DE SR345 coding sequence.
XX
XX Gene isolation; membrane-bound protein; fusion protein; drug production;
KW antigen-binding cell; secretable functional protein; antigenic protein;
KW protein isolation; diagnosis; SR345 protein; ss.
XX
XX Homo sapiens.
XX
XX WO9960113-A1.
XX
XX 25-NOV-1999.
XX
XX 30-APR-1999; 99WO-JP02341.
XX
XX 20-MAY-1998; 98JP-0138652.
XX
XX 01-OCT-1998; 98JP-0279876.
XX
XX (CHUS ) CHUGAI SEIYAKU KK.
XX
XX Tsuchiya M, Saito M, Ohtomo T;
XX
XX WPI; 2000-039382/03.
XX
XX P-PSDB; AAY55071.
XX
XX Efficient and selective isolation of a gene encoding membrane protein
PT with low or no antigenic binding activity, for diagnosis, study of, and
PT production of drugs treating abnormal functions of the protein -
XX
XX Example 1; Page 49-52; 120pp; Japanese.
XX
XX This sequence encodes the SR345 protein.
CC The invention relates to a method for isolating a gene encoding a
CC membrane-bound protein, comprising introducing a vector into a cell,
CC contacting an antigen with the cell expressing the fused protein encoded
```

CC by the vector on its surface to select an antigen-binding cell, and  
CC isolating the cDNA. The vector contains DNA encoding a secretable  
CC functional protein with antigenicity and binding affinity, and a cDNA  
CC ligated to DNA downstream of the 3' end of the coding sequence. The  
CC method can be used to isolate a membrane-bound protein for diagnosis and  
CC study. It can also be used for producing drugs treating abnormal  
CC functions of the protein. Such a technique is efficient and selective,  
CC which is different from the prior-art transmembrane trap (TMT) method  
CC wherein an epitope recognised by an antibody is carried in a fused  
CC protein.

XX  
SQ Sequence 1035 BP; 210 A; 311 C; 315 G; 199 T; 0 other;

Query Match 23.1%; Score 35.8; DB 21; Length 1035;  
Best Local Similarity 54.1%; Pred. No. 0.16;  
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 10 ATACCTGCTCTTAATACGGCAGCTATTGAGCATGTAGAAGTTGCTGGACCTCCAGCACAC 69  
|||||  
DB 768 ATATCGGGCTGAACGGTCAAGACATTCACAAACATGGATGGTCAAGGACCTCCAGCATCA 827  
|||||  
QY 70 CCCAGGCCCCAGAGAAGTGGGGCTCTCTGTGCACCAAGTTTACCACAATATACAGGA 129  
|||||  
DB 828 CTGTGTCATCCACGAGCGCTGAGGCGCTGAGGCACGTGGTGCAGCTTCGTGCCCAGGA 887  
|||||  
QY 130 GAAATAAGTGAATG 144  
|||||  
DB 888 GGAGTTCGGGCAAGG 902

RESULT 6  
AAV04440  
ID AAV04440 standard; DNA; 1074 BP.  
XX  
AC AAV04440;  
XX  
DT 27-APR-1998 (first entry)  
XX  
DE Interleukin 6 receptor DNA.  
XX  
KW Interleukin 6 receptor; human; screening;  
KW combinatorial library; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9737220-A1.  
XX  
PD 09-OCT-1997.  
XX  
PF 02-APR-1997; 97WO-US05821.  
XX  
PR 03-APR-1996; 96US-0627151.  
XX  
PA (CHUG-) CHUGAI BIOPHARMACEUTICALS INC.  
XX  
PI Becherer KA, Brown SJ, Spinella DG;  
XX  
DR WPI; 1997-503233/46.  
XX  
PT Screening of compounds for ability to bind specific molecules -  
PT using a chimeric protein in which the specific molecule is fused to  
PT an immunoglobulin chain  
XX  
PS Example 5; Pages 39-40; 80pp; English.  
XX  
CC The present sequence was used in the development of a novel method  
CC of screening compounds for ability to bind a specific molecule. The  
CC method comprises contacting one or more compounds with a chimeric  
CC protein containing at least two domains, the 1st comprising a  
CC portion of the specific molecule, and the 2nd an immunoglobulin  
CC chain portion having one or more epitopes and/or immunoglobulin  
CC regions recognising an epitope. A binding partner complex between  
CC the chimeric protein and compound(s) is formed, separated out and

CC contacted with a (in)directly labelled secondary molecule which  
CC binds the 2nd domain, and the label detected. Using the chimeric  
CC proteins, compounds can be rapidly screened for binding to an  
CC antigen, antibody, enzyme, enzyme substrate, receptor or ligand,  
CC e.g. in biomedical research and drug development. The method is  
CC especially intended for screening combinatorial libraries, but is  
CC also useful in screening bacterial/phage lysates, assays requiring  
CC specific binding partner interaction and obtaining binding  
CC analogues of a compound.

XX  
SQ Sequence 1074 BP; 223 A; 321 C; 324 G; 206 T; 0 other;

Query Match 23.1%; Score 35.8; DB 18; Length 1074;  
Best Local Similarity 54.1%; Pred. No. 0.16;  
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 10 ATACCTGCTCTTAATACGGCAGCTATTGAGCATGTAGAAGTTGCTGGACCTCCAGCACAC 69  
|||||  
DB 768 ATATCGGGCTGAACGGTCAAGACATTCACAAACATGGATGGTCAAGGACCTCCAGCATCA 827  
|||||  
QY 70 CCCAGGCCCCAGAGAAGTGGGGCTCTCTGTGCACCAAGTTTACCACAATATACAGGA 129  
|||||  
DB 828 CTGTGTCATCCACGAGCGCTGAGGCGCTGAGGCACGTGGTGCAGCTTCGTGCCCAGGA 887  
|||||  
QY 130 GAAATAAGTGAATG 144  
|||||  
DB 888 GGAGTTCGGGCAAGG 902

RESULT 7  
AAZ09202  
ID AAZ09202 standard; DNA; 1260 BP.  
XX  
AC AAZ09202;  
XX  
DT 19-OCT-1999 (first entry)  
XX  
DE Human IL-6 receptor/IL-6 fusion protein DNA.  
XX  
KW IL-6; interleukin-6; receptor; human; fusion protein; bone marrow;  
KW stem cell; platelet; reduced antigenicity; ds.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
FH Key  
FT CDS  
FT Location/Qualifiers  
FT 1..1260  
FT /\*tag= a  
FT /product= "IL-6 receptor/IL-6 fusion protein"  
FT /note= "No start codon given"  
XX  
PN JP11196867-A.  
XX  
PD 27-JUL-1999.  
XX  
PF 09-JAN-1998; 98JP-0002921.  
XX  
PR 09-JAN-1998; 98JP-0002921.  
XX  
PA (TOYJ ) TOSOH CORP.  
XX  
DR WPI; 1999-496648/42.  
DR P-PSDB; AAY30938.  
XX  
PT New interleukin-6 receptor-interleukin-6 fused protein and gene -  
PT used for growth of bone marrow stem cells and platelets  
XX  
PS Example 1; Page 5; 8pp; Japanese.  
XX  
CC This invention describes a novel gene which encodes a fusion protein of  
CC interleukin-6 (IL-6) receptor and bound with a gene sequence encoding  
CC for IL-6 at the downstream of IL-6 receptor gene. The gene and its  
CC encoding protein has applications for the growth of bone marrow stem



Query Match 23.1%; Score 35.8; DB 21; Length 1545;  
Best Local Similarity 54.1%; Pred. No. 0.19;  
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 10 ATACCTGTCTTATAACGGCAGTCATTGAGCATGTAGAAAGTTGCTGGACCTCCAGCACAC 69  
Db 711 ATATCGGGCTGAACGGCTCAAGACATTCACACATGGATGGTCAAGGACCTCCAGCATCA 770  
QY 70 CCAGGCCCCCAAGAAAGTGGGGCCTCTGTGTGCACCAAGGTTTACCACAATATACAGGA 129  
Db 771 CTGTGTTCATCCAGCAGCGCCTGGAGCGCCTGAGGCACGTGGTGCACGCTTCGTGCCCAGGA 830  
QY 130 GAAATAAGTGAATG 144  
Db 831 GGAGTTCGGGCAAGG 845

RESULT 10  
AAT97848  
ID AAT97848 standard; DNA; 1627 BP.  
AC AAT97848;  
DT 25-MAR-1998 (first entry)  
DE Human fusion polypeptide H-IL-6 DNA containing a 18 amino acid linker.  
KW Interleukin-6; IL-6; interleukin-6 receptor; IL-6R; ligand; conjugate;  
KW protein interaction; therapeutic; antagonist; ss.  
OS Synthetic.  
OS Homo sapiens  
XX  
XX  
FH Key Location/Qualifiers  
CDS 34..1611  
FT /\*tag= a  
FT sig\_peptide 34..90  
FT /\*tag= b  
FT mat\_peptide 91..1608  
FT /\*tag= c  
FT /product= H-IL-6  
FT /note= "fusion polypeptide"

XX WO9732891-A2.  
PN  
PD 12-SEP-1997.  
XX  
PF 07-MAR-1997; 97WO-DE00458.  
XX  
PR 07-MAR-1996; 96DE-4008813.  
XX  
PA (ANGE-) ANGEWANDTE GENTECHNOLOGIE SYSTEME GMBH.  
XX  
PI Rose-John S;  
XX  
XX WPI; 1997-470536/43.  
DR P-PSDB; AAW36846.  
XX  
XX Conjugate of two peptide(s) with mutual affinity connected by a  
PT linker - used to modulate interactions between proteins, e.g. for ex  
PT vivo expansion of human stem cells  
XX  
PS Claim 12; Fig 1; 19pp; German.  
XX  
XX This sequence encodes the fusion polypeptide H-IL-6 which contains an  
CC 18 amino acid linker which joins the carboxy terminus of human  
CC interleukin-6 receptor (IL-6R) with the amino terminus of human  
CC interleukin-6 (IL-6). Such conjugates could be used to modulate  
CC interactions between proteins, particularly to overcome interrupted  
CC interactions caused by an incomplete interleukin-6 (IL-6) receptor. These  
CC constructs derived from IL-6 and its receptor, can also be used for ex  
CC vivo expansion of human stem cells, and as a therapeutic IL-6 receptor

CC antagonist.  
XX  
SQ Sequence 1627 BP; 388 A; 450 C; 481 G; 308 T; 0 other;  
Query Match 23.1%; Score 35.8; DB 18; Length 1627;  
Best Local Similarity 54.1%; Pred. No. 0.19; Mismatches 62; Indels 0; Gaps 0;  
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 10 ATACCTGTCTTATAACGGCAGTCATTGAGCATGTAGAAAGTTGCTGGACCTCCAGCACAC 69  
Db 801 ATATCGGGCTGAACGGCTCAAGACATTCACACATGGATGGTCAAGGACCTCCAGCATCA 860  
QY 70 CCAGGCCCCCAAGAAAGTGGGGCCTCTGTGTGCACCAAGGTTTACCACAATATACAGGA 129  
Db 861 CTGTGTTCATCCAGCAGCGCCTGGAGCGCCTGAGGCACGTGGTGCACGCTTCGTGCCCAGGA 920  
QY 130 GAAATAAGTGAATG 144  
Db 921 GGAGTTCGGGCAAGG 935

RESULT 11  
AAN90340  
ID AAN90340 standard; cDNA; 2061 BP.  
XX  
AC AAN90340;  
DT 31-MAR-1992 (first entry)  
DE Sequence encoding a receptor protein for human B cell stimulating  
DE factor-2 (BSF2 receptor).  
XX  
KW B cell; immune disorder; therapy; diagnosis; prophylaxis; ss.  
XX Homo sapiens.  
XX  
FH Key Location/Qualifiers  
CDS 247..1753  
FT /\*tag= a  
FT /note= "SQ claimed"  
XX  
PN EP325474-A.  
XX  
PD 26-JUL-1989.  
XX  
PF 20-JAN-1989; 89EP-0300536.  
XX  
PR 14-JAN-1989; 89JP-0017461.  
PR 22-JAN-1988; 88JP-0012387.  
PR 25-JAN-1988; 88JP-0012599.  
PR 04-AUG-1988; 88JP-0194885.  
PR 20-JAN-1989; 89JP-0009774.  
XX  
PA (KISH/) KISHIMOTO T.  
XX  
PI Kishimoto T;  
XX  
XX WPI; 1989-214667/30.  
DR P-PSDB; AAP90284.  
XX  
XX Receptor protein for human B cell stimulating factor-2 - obtd. by  
PT recombinant DNA techniques and used as diagnostic, prophylactic or  
PT therapeutic agent  
XX  
PS Disclosure; Fig 3; 63pp; English.  
XX  
XX The cDNA in AAN90340 was derived from monocytic cell line U937.  
CC Isolated BSF2 receptor and DNA encoding it are claimed, as are  
CC (b) expression vectors; (c) host organisms; (d) antibodies; and  
CC (e) hybridomas.  
XX  
XX Sequence 2061 BP; 418 A; 631 C; 621 G; 391 T; 0 other;  
SQ



Db	1074	CTGTGTCATCCAGCAGCGCTGGAGGGCGCTGAGGCACTGCTGTCAGCTTCGTGCCAGGA	1133
Qy	130	GAATAAGTGAATG	144
Db	1134	GGAGTTCGGCAAGG	1148
RESULT 14			
AAT31441			
ID	AAT31441	standard; cDNA; 3319 BP.	
XX	AAT31441;		
XX	28-NOV-1996	(first entry)	
XX	Interleukin-6	receptor coding sequence.	
XX	Interleukin-6; IL; receptor; antisense oligonucleotide; inhibition;		
KW	gene expression; kidney tumour; myeloma; Kaposi's sarcoma; psoriasis;		
KW	rheumatoid arthritis; endotoxic shock; ss.		
XX	Homo sapiens.		
OS			
XX	Key	Location/Qualifiers	
FH	CDS	438..1844	
FT		/*tag= a	
FT		/product= Interleukin-6 receptor.	
XX			
PN	W09618416-A1.		
XX			
PD	20-JUN-1996.		
XX			
PF	15-DEC-1995; 95WO-JP02587.		
XX			
PR	18-AUG-1995; 95JP-0210739.		
PR	16-DEC-1994; 94JP-0913167.		
XX			
PA	(CHUS ) CHUGAI SEIVAKU KK.		
XX			
PI	Koishibara Y, Kuromaru K;		
XX			
DR	WPI; 1996-300392/30.		
DR	P-PSDB; AAR98364.		
XX			
PT	Anti-sense oligo:nucleotide inhibitor against human IL-6R expression		
PT	- for treatment of e.g. tumours, cancers, rheumatoid arthritis,		
PT	psoriasis, endotoxic shock, etc.		
XX			
PS	Claim 2; Page 17-21; 32pp; Japanese.		
XX			
CC	Antisense oligonucleotides may be used to inhibit the expression of		
CC	the interleukin-6 receptor. Inhibition of expression of the		
CC	IL-6 receptor is useful in the treatment of kidney tumours, myeloma,		
CC	Kaposi's sarcoma, rheumatoid arthritis, psoriasis and endotoxic		
CC	shock. The antisense oligonucleotides are administered at a dosage		
CC	of 0.1-100mg/kg, pref. 0.1-50 mg/kg.		
XX			
SQ	Sequence 3319 BP; 735 A; 937 C; 960 G; 687 T; 0 other;		
Query Match	23.1%; Score 35.8; DB 17; Length 3319;		
Best Local Similarity	54.1%; Pred. No. 0.25;		
Matches	73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;		
Qy	10	ATACCTGCTTAATACGGCAGTCATTGAGCATGTAGAAAGTTGCTGGACCTCCAGCAC	69
Db	1205	ATATCGGGCTGAACGGTCAAGACATTCACAAACATGGTATGGTCAAGGACCTCCAGCATCA	1264
Qy	70	CCAGGGCCCCAGAAAGTGGGGCTCTCGTGCACCAGGTTTACCACAATATACAGGA	129
Db	1265	CTGTGTCATCCAGCAGCGCTGGAGGGCGCTGAGGCACTGCTGTCAGCTTCGTGCCAGGA	1324
Qy	130	GAATAAGTGAATG	144

Db	1325	GGAGTTCGGCAAGG	1339	
RESULT 15				
AAV60295				
ID	AAV60295	standard; DNA; 3319 BP.		
XX	AC			
XX	AAV60295;			
XX	02-FEB-1999	(first entry)		
XX	Human interleukin-6 receptor cDNA.			
DE	Interleukin-6 receptor; human; hepatitis B virus; HBV; infection; therapy; ss.			
XX	XX			
XX	Homo sapiens.			
XX	XX			
XX	W09835694-A2.			
XX	XX			
XX	20-AUG-1998.			
XX	XX			
PF	10-FEB-1998;	98WO-US08898.		
XX	XX			
PR	11-FEB-1997;	97US-0795473.		
XX	XX			
PA	(DAVI/) DAVIDSON C M.			
PA	(HADA-) HADASIT MEDICAL RES SERVICES & DEV.			
XX	XX			
DR	WPT; 1998-520755/44.			
XX	XX			
PT	Treatment of hepatitis B virus infection - using a soluble active agent which prevents interaction of HBV with hepatocytes mediated by human interleukin 6			
PT	Disclosure; Fig 6a-b; 51pp; English.			
XX	XX			
CC	This is a previously reported nucleotide sequence for human interleukin-6 (hIL-6) receptor mRNA. The invention relates to the finding that hIL-6 is essential for hepatitis B virus (HBV) infection. The invention provides a pharmaceutical composition for the treatment of HBV infection, comprising a soluble active agent that interacts with at least one of the binding sites between hIL-6 and the pS1 region of HBV and between hIL-6 and hepatocytes and other HBV-permissive cells. The active agent competitively binds to at least one of these sites and thereby prevents hIL-6-mediated HBV infection of hepatocytes and other HBV-permissive cells. The soluble active agent is selected from glycoprotein 80 (gp80), having receptor sites which interact with hIL-6, soluble glycoprotein 130 (gp130) having receptor sites which interact with hIL-6, hIL-6 derived peptide Lys41-Ala56, hIL-6 derived peptide Gly77-Glu95, hIL-6 derived peptide Gln153-His165, a combined 1 and 2 hIL-6 mutant (mhIL-6 1+2), and mhIL-6 1+2 substituted with Phe171 to Leu and Ser177 to Arg, and mixtures of any of these.			
XX	Sequence 3319 BP; 735 A; 936 C; 961 G; 687 T; 0 other;			
XX	XX			
Query Match	23.1%;	Score 35.8; DB 19; Length 3319;		
Best Local Similarity	54.1%;	Pred. No. 0.25;		
Matches	73; Conservative	0; Mismatches 62; Indels 0; Gaps		
Qy	10	ATACCTGCTTAATAACGGCAGTCATTGAGCATGTAGAGTTGCTGGACCTCCAGCACAC	69	
Db	1205	ATAATCGGGCTGAACGGGTCAAGACATTCACACATGGATGGTCAAGGACCTCCAGCATCA	1264	
Qy	70	CCCAAGGCCCCAGAAAGTGGGGCCCTCCTGGTGCCACCAAGGTTTACCACAATATACAGGA	129	
Db	1265	CTGTGTATCCACGACGCCTGGAGCGCCTGAGGCACGCTGGTGCACTTCGTGCCACGGA	1394	
Qy	130	GAATAAGTGAATG	144	
Db	1325	GGAGTTCGGCAAGG	1339	

Search completed: February 19, 2003, 22:56:26  
Job time : 61.6445 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:16:26 ; Search time 494.184 Seconds  
(without alignments)  
9128.055 Million cell updates/sec

Title: US-09-997-610-1\_COPY\_2\_156

Perfect score: 155

Sequence: 1 atgtgtgtctactctgtctt.....agtgaatgacaaaatgcc 155

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	111	71.6	145880	9	HS302D9	Z82198 Human DNA s
2	40.6	26.2	5851	3	AF282902	AF282902 Hydra vul
3	39.2	25.3	87100	9	HS43814	Z97635 Human DNA s
4	38.2	24.6	1395	5	DRE318214	AJ318214 Danio rer
5	38	24.5	287	3	OOS420010	AJ420010 Osteragi
6	37.8	24.4	9448	3	CEU22327	U22327 Caenorhabd
7	37.8	24.4	9541	3	CECOLA21V	Z22964 C. elegans a
8	37.8	24.4	34671	3	U53342	U53342 Caenorhabd
9	37.2	24.0	123789	2	AC110674	AC110674 Tetraodon
10	36.8	23.7	3612	10	MWCOLA4	X67777 Mouse mRNa
11	36.8	23.7	6512	6	AX306181	AX306181 Sequence
12	36.8	23.7	6512	10	MUSCOLIA4A	J04694 Mus muscu
13	36.8	23.7	7765	10	AF169387	AF169387 Mus muscu
14	36.6	23.6	694	3	HSNCOL1	X61045 Hydra N-COL
15	36.6	23.6	95492	2	AC096135	AC096135 Rattus no
16	36.2	23.4	36532	3	CEF57B1	Z78064 Caenorhabd
17	36.2	23.4	269619	3	CEY51H4A	AL132952 Caenorhab
18	35.8	23.1	975	6	E64808	E64808 Yeast trans
19	35.8	23.1	1074	6	AR031384	AR031384 Sequence
20	35.8	23.1	1074	6	BD009752	BD009752 Composi
21	35.8	23.1	1260	6	E28089	E28089 Novel IL-6
22	35.8	23.1	1486	6	ARI45506	ARI45506 Sequence
23	35.8	23.1	1486	9	HS16REC	X58298 Human mRNa
24	35.8	23.1	1545	6	E35612	E35612 IL-6 recept
25	35.8	23.1	1612	6	A93715	A93715 Sequence 2
26	35.8	23.1	1627	6	A93714	A93714 Sequence 1
27	35.8	23.1	2066	6	E04823	E04823 cDNA encodi
28	35.8	23.1	2087	6	E02673	E02673 cDNA encodi
29	35.8	23.1	3319	6	ARI45505	ARI45505 Sequence
30	35.8	23.1	3319	6	AX399205	AX399205 Sequence
31	35.8	23.1	3319	6	E12979	E12979 cDNA encodi
32	35.8	23.1	3319	9	HSIL6R	X12830 Human mRNa
33	35.6	23.0	3734	14	HVSX99519	X99519 Herpesvirus
34	35.6	23.0	11942	1	AE010554	AE010554 Fusobacte
35	35.4	22.8	130619	2	AC120069	AC120069 Rattus no
36	35.2	22.7	1234	6	AX331924	AX331924 Sequence
37	35.2	22.7	1234	9	HSU52521	U52521 Human arfap
38	35.2	22.7	1993	3	BMCOLGMR	Z30348 B.mori mRNa
39	35	22.6	1973	9	HS10ALCOL	X65120 H.sapiens C
40	35	22.6	1976	10	RN005395	AJ005395 Rattus no
41	35	22.6	2217	10	RNPRO1C	X70369 R.norvegicu
42	35	22.6	3215	9	HSCOLX3	X72580 Homo sapien
43	35	22.6	3226	6	AX33243	AX33243 Sequence
44	35	22.6	3226	9	HSCOLA1X	X60382 H.sapiens C
45	35	22.6	3226	11	G28608	G28608 human STS S

ALIGNMENTS

RESULT 1  
HS302D9  
LOCUS Human DNA sequence from clone RPI-302D9 on chromosome 22 Contains  
DEFINITION HS302D9 145880 bp DNA linear PRI 12-DEC-1999  
GSSs, complete sequence.  
ACCESSION Z82198  
VERSION Z82198.2 GI:6572207  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 145880)  
AUTHORS Bridgeman,A.  
TITLE Direct Submission

JOURNAL

Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk) Clone requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
On Dec 13, 1999 this sequence version replaced gi:3164067.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.  
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>  
RPI-302D9 is from the library RPCI-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pcypac2>

This sequence is the entire insert of clone RPI-302D9 The true left end of clone CTA-282F2 is at 69682 in this sequence. The true right end of clone CTA-415G2 is at 55167 in this sequence.

FEATURES

source

1..145880  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="22"  
/clone="RPI-302D9"  
/clone\_lib="RPCI-1"

repeat\_region 188..245  
/note="MER3 repeat: matches 144..209 of consensus"  
repeat\_region 246..571  
/note="AluX repeat: matches 1..312 of consensus"  
repeat\_region 572..759  
/note="MER3 repeat: matches 1..144 of consensus"  
repeat\_region 783..933  
/note="MERA repeat: matches 26..187 of consensus"  
repeat\_region 1033..1336  
/note="AluSp repeat: matches 1..299 of consensus"  
repeat\_region 1450..1583  
/note="MIR repeat: matches 24..160 of consensus"  
repeat\_region 1687..1752  
/note="L2 repeat: matches 2593..2661 of consensus"  
repeat\_region 2350..2660  
/note="AluSc repeat: matches 3..309 of consensus"  
repeat\_region 2684..2981  
/note="AluSq repeat: matches 2..300 of consensus"  
repeat\_region 3323..3343  
/note="MLTIE repeat: matches 116..136 of consensus"  
repeat\_region 3344..3652  
/note="AluY repeat: matches 1..309 of consensus"  
repeat\_region 3653..3928  
/note="MLTIE repeat: matches 136..359 of consensus"  
repeat\_region 3929..4278  
/note="THEB repeat: matches 3..364 of consensus"  
repeat\_region 4279..4485  
/note="MLTIE repeat: matches 359..568 of consensus"  
repeat\_region 5073..5176  
/note="52 copies 2 mer ct 78 conserved"  
repeat\_region 5181..5491  
/note="AluJb repeat: matches 1..311 of consensus"  
repeat\_region 6369..6485  
/note="L2 repeat: matches 2579..2705 of consensus"

repeat\_region 6647..6685  
/note="MADE1 repeat: matches 1..23 of consensus"  
repeat\_region 6686..6987  
/note="AluX repeat: matches 1..302 of consensus"  
repeat\_region 6988..7036  
/note="MADE1 repeat: matches 23..77 of consensus"  
repeat\_region 7482..7754  
/note="AluJb repeat: matches 9..290 of consensus"  
repeat\_region 7775..8060  
/note="AluJo repeat: matches 1..295 of consensus"  
repeat\_region 8414..8551  
/note="L2 repeat: matches 2553..2706 of consensus"  
repeat\_region 8914..9030  
/note="MIR repeat: matches 147..262 of consensus"  
repeat\_region 9110..9280  
/note="MIR repeat: matches 91..262 of consensus"  
repeat\_region 9283..9412  
/note="MIR repeat: matches 15..144 of consensus"  
repeat\_region 9521..9679  
/note="FAM repeat: matches 3..161 of consensus"  
repeat\_region 9820..10225  
/note="MSTB repeat: matches 2..425 of consensus"  
misc\_feature complement(10179..10678)  
/note="match: GSS: Em:B56592"  
misc\_feature complement(10204..10728)  
/note="match: GSS: Em:AQ701486"  
misc\_feature complement(10249..10706)  
/note="match: GSS: Em:AQ225495"  
repeat\_region 10312..10383  
/note="MIR repeat: matches 79..150 of consensus"  
misc\_feature 10718..11310  
/note="match: GSS: Em:B14024"  
misc\_feature 10784..11201  
/note="match: GSS: Em:B43656"  
repeat\_region 11838..11946  
/note="MIR repeat: matches 20..137 of consensus"  
repeat\_region 12174..12445  
/note="L2 repeat: matches 1988..2275 of consensus"  
repeat\_region 12444..12642  
/note="MIR repeat: matches 63..241 of consensus"  
misc\_feature 13017..13369  
/note="match: STS: Em:G49301"  
repeat\_region 13331..13397  
/note="MIR repeat: matches 174..244 of consensus"  
repeat\_region 13398..13698  
/note="AluSp repeat: matches 1..302 of consensus"  
repeat\_region 13699..13810  
/note="MIR repeat: matches 76..174 of consensus"  
repeat\_region 13806..13919  
/note="MIR repeat: matches 77..189 of consensus"  
repeat\_region 13945..14060  
/note="MIR repeat: matches 24..142 of consensus"  
repeat\_region 14061..14367  
/note="AluY repeat: matches 1..301 of consensus"  
repeat\_region 14368..14452  
/note="MIR repeat: matches 141..225 of consensus"  
repeat\_region 14589..14679  
/note="MIR repeat: matches 173..262 of consensus"  
misc\_feature 14597..15201  
/note="match: GSS: Em:AQ553482"  
misc\_feature 14616..15060  
/note="match: GSS: Em:AQ370601"  
repeat\_region 14868..15040  
/note="MIR repeat: matches 49..233 of consensus"  
repeat\_region 15071..15188  
/note="L2 repeat: matches 2112..2239 of consensus"  
repeat\_region 15304..15399  
/note="MLTIB repeat: matches 1..99 of consensus"  
repeat\_region 15490..15662  
/note="AluSg1 repeat: matches 2..114 of consensus"  
repeat\_region 15669..15727  
/note="MLTIB repeat: matches 119..178 of consensus"  
repeat\_region 15728..16027

repeat\_region /note="AluSc repeat: matches 1. .299 of consensus"  
16028. .16245  
repeat\_region /note="MLT1B repeat: matches 178. .390 of consensus"  
16546. .16854  
repeat\_region /note="AluY repeat: matches 1. .300 of consensus"  
18296. .18323  
repeat\_region /note="MSTA repeat: matches 2. .29 of consensus"  
18324. .18392  
repeat\_region /note="MER66-internal repeat: matches 4919. .4993 of  
consensus"  
18393. .18712  
repeat\_region /note="AluJb repeat: matches 1. .311 of consensus"  
18713. .19133  
repeat\_region /note="MER66-internal repeat: matches 4548. .4919 of  
consensus"  
complement(18872. .19230)  
misc\_feature /note="match: GSS: Em:AQ005063"  
19251. .19719  
misc\_feature /note="match: GSS: Em:B14179"  
19537. .20290  
repeat\_region /note="HERVH21 repeat: matches 4657. .5784 of consensus"  
20317. .20382  
repeat\_region /note="33 copies 2 mer ta 68 conserved"  
20313. .20666  
repeat\_region /note="77 copies 2 mer tt 70 conserved"  
20682. .21008  
repeat\_region /note="AluSg1 repeat: matches 1. .306 of consensus"  
21239. .21553  
repeat\_region /note="HUPRS-P3 repeat: matches 4410. .4713 of consensus"  
21882. .22254  
repeat\_region /note="THE1B repeat: matches 1. .364 of consensus"  
22302. .22537  
repeat\_region /note="MER66-internal repeat: matches 2186. .2417 of  
consensus"  
22538. .22850  
repeat\_region /note="AluSp repeat: matches 1. .313 of consensus"  
22851. .23801  
repeat\_region /note="MER66-internal repeat: matches 1210. .2186 of  
consensus"  
23905. .23989  
repeat\_region /note="MER66-internal repeat: matches 3017. .3102 of  
consensus"

Query Match 71.0%; Score 111; DB 9; Length 145880;  
Best Local Similarity 100.0%; Pred. No. 3e-23;  
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATAGTGTACATCTGCTTAAATACCGCAGTCTTATGAGCATGTAGAGTTGCTGGACCT 60  
Db 34786 ATAGTGTACATCTGCTTAAATACCGCAGTCTTATGAGCATGTAGAGTTGCTGGACCT 34845  
Qy 61 CCAGCACACCCAGGCCCCCAGAGAGAGTGGGGCTCTCTGTCGCACACAGGT 111  
Db 34846 CCAGCACACCCAGGCCCCCAGAGAGAGTGGGGCTCTCTGTCGCACACAGGT 34896

RESULT 2  
AF282902 5851 bp mRNA linear INV 17-DEC-2000  
LOCUS Hydra vulgaris type IV collagen alpha 1 chain precursor, mRNA,  
DEFINITION complete cds.

ACCESSION AF282902  
VERSION AF282902.1 GI:11875611

KEYWORDS Hydra vulgaris.  
SOURCE Hydra vulgaris.  
ORGANISM Hydra vulgaris.  
Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;  
Hydridae; Hydra.  
REFERENCE 1 (bases 1 to 5851)  
AUTHORS Fowler,S.J., Jose,S., Zhang,X., Deutzmann,R., Sarraz,M.P. Jr. and  
Boot-Handford,R.P.  
TITLE Characterization of hydra type IV collagen. Type IV collagen is  
essential for head regeneration and its expression is up-regulated  
upon exposure to glucose

J. Biol. Chem. 275 (50), 39589-39599 (2000)  
JOURNAL 20564332  
MEDLINE 10956657  
PUBMED 2 (bases 1 to 5851)  
REFERENCE Fowler,S.J. and Boot-Handford,R.P.  
AUTHORS Direct Submission  
TITLE Submitted (27-JUN-2000) School of Biological Sciences, University  
JOURNAL of Manchester, Oxford Road, Manchester M13 9PT, UK  
FEATURES Location/Qualifiers  
source 1. 5851  
/organism="Hydra vulgaris"  
/db\_xref="taxon:5087"  
157. .5328  
/product="type IV collagen alpha 1 chain precursor"  
/protein\_id="AAG40729.1"  
/db\_xref="GI:11875612"  
/translation="MNHTTKWTIWIIVLFTLHSLCLVTLQCGGGCCNQCINVCVQKQK  
DRGVSGLPGFKPIGEPFGPGEPAGPGKGGGAPGLQGVGARGKVGPPGPTGI  
PGSHRPGDEGRQEGPLPCNGTKGDIIGPPGRKDVQVQGLPGPIGRPPGDS  
TATLKQKQEGPGPKSQDGPAGDPGTGDKGESCAGRMGPGQLRGKARGDSDNITI  
FGERGDKDILGPPGRDNCSSVTGELTINIOGPKGEOCKGDKGKGGOPGOA  
GEAGDQKQKGEKDKGEGISNGPSIGQGEKGDIGATNAATGKBPESDGLSGSQKQK  
DHGTKGSEGPVDRGKQGESITGPPQIGEKGRQKGEKGPPEGPPNGEVGDT  
DPPGLKGLKEIGMTGSGEPGTGAECMKGPAGQGTGKESGSRPGSVQGS  
PGMDNPGQIGPPGEPGIPGPPGPNQVYNDLGETVLPFPDQDQPGQIQGVGDP  
QSGPIRGDKGACKSCPSGPGQGVAGDGLPCANGKDKGDKGKPGLEGSVGP  
GESLTPQPGIPEGKDTGAKNGKSGDRIVOGERGEKTDGMPGSGSIGRKGEO  
GVGPDGTDGPGQGLQGGDRGAPKSGIIPNNGEKEKGNKGDGLRGRGGER  
DDEPGKSGQPMKPGPKGPEFFSTALKQGGDKDGVSEGGPPGAGGVGNQGF  
KEPGLRAGSGEKEGKADAGTPIAGKNGENGKDDQPKGDKGKGLQPIGISGP  
IKEPGLKGFPGPLGDAPKSVKPGEPGRDGDGDPKAGPEPTGEGSRDEKGE  
IGPQPGDKDQGVSGVIGPKGTGAIGKGTGGDLKGRDGIKQAGKEKEKGN  
SVGLPGDKGKSGPGKDKGISGKQKPGPKGKDVGTGPKDGLPGNPGEPKNG  
EKGVTPGLDGEKGEPCYSCQKAKGTGVRGKPGATSNVWPKGNRGEKGEPC  
KQPPGAEQSGEGERGIGTDKGTGDTGIGLPKGERGERGERGKQDQGPPIG  
KQVPEGPAGSIGERTIGNDGPKGSGPPIKGEPSVGSIGPKSKGEGGKGER  
ESTLKNPGEPRGPTGPPGAGKYLIDLEVLGKEKPTKBPEDGPKGDKGIDG  
GTSQGTGKGEKGVKGEQGLGPKGLGVRVLAGSKDGIKGVIGKPGTGD  
KKEGLQGFKGTGPGTQKGEKGEKGLGKGEKAGKMGKMPDAGATIGKEK  
DREKESAGQRYKGEPSQSLTINPCKAGEQDRIKGEKNGIKGPNRGEKDG  
SIGEPKSGQKGTGLDGPKEKQPAIGTGPKEKQDQGPQPGQPGQDGLGKGG  
DIPPGETDIPGQIKGEFAGIAGTGPGLQGLGKQDQGPQDQGPQDQGPQGR  
SGDGFYGLVKHSQISKVPSCAGMTMWEGYFLYAQNERRAFQDGLQPGSLKR  
FSTWPLFQIDQNKVVASRNDYSFSLSTAEKPKAESGADLENYISRCIVCAPSH  
VLAVHSELDKPCPDGWNLTGFSFLMYNSAGAOGSGLLSGSGCLDFRNPYI  
ECHRGTCTWYGYPTLSFWLSTIGESNMFQVFPKFEILLRNKARYSRCAVCKSVP"  
157. .228  
sig\_peptide 229. .276  
misc\_feature /note="Region: non-collagenous domain"  
277. .4641  
misc\_feature /note="Region: contains collagenous domain"  
526. .528  
/note="N-glycosylation"  
/evidence=not\_experimental  
4642. .5325  
misc\_feature /note="Region: non-collagenous domain"  
/note="Region: non-collagenous domain"  
BASE COUNT 1953 a 942 c 1638 g 1318 t  
ORIGIN

Query Match 26.2%; Score 40.6; DB 3; Length 5851;  
Best Local Similarity 64.2%; Pred. No. 0.12;  
Matches 61; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 55 GGACCTCCAGCACACCCAGGCCCCCAGAGAGTGGGGCTCTCTGTCGCACACAGTTTA 114  
Db 2065 GGACCTCAAGGACTTCAAGGCTCTCAAGGTGATCTGTGCTCTCTGGTAAATCAGGTATA 2124  
Qy 115 CCACATAATACAGGAGAAATAGTGAATGACAAA 149  
Db 2125 CCAGGCATACCGAGAGAAAGGTGAAAAGGAAA 2159

```
RESULT 3
HS438L4/c
LOCUS
DEFINITION
  HS438L4 87100 bp DNA linear PRI 23-NOV-1999
  Human DNA sequence from clone 438L4 on chromosome 1p36.2-36.3
  Contains ESTs, STS and GSSs, complete sequence.
ACCESSION
  Z97635
VERSION
  HTG.
KEYWORDS
  Z97635.10 GI:4835272
SOURCE
  Homo sapiens.
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  1 (bases 1 to 87100)
  Direct Submission
  Submitted (08-JUN-1999) Sanger Centre, Hinxton, Cambridgeshire,
  CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
  requests: clonerequest@sanger.ac.uk
COMMENT
  On May 14, 1999 this sequence version replaced gi:4678473.
  During sequence assembly data is compared from overlapping clones.
  Where differences are found these are annotated as variations
  together with a note of the overlapping clone name. Note that the
  variation annotation may not be found in the sequence submission
  corresponding to the overlapping clone, as we submit sequences with
  only a small overlap as described above.
  The following abbreviations are used to associate primary accession
  numbers given in the feature table with their source databases:
  Em: EMBL; Sw: SWISSPROT; Tr: TREMBL
  IMPORTANT: This sequence is not the entire insert of clone 438L4.
  It may be shorter because we only sequence overlapping sections
  once, or longer because we arrange for a small overlap between
  neighbouring submissions.
  The true left end of clone 505B13 is at 87001 in this sequence.
  This sequence has been finished according to sequence map criteria
  as follows. An attempt is made to resolve all sequencing problems,
  such as compressions and repeats, but not necessarily within known
  annotated human repeat sequence elements (e.g. Alu). Where the
  sequence is ambiguous, there is an annotation using the 'unsure'
  feature key.
  This sequence was generated from part of bacterial clone contigs of
  human chromosome 1, constructed by the Sanger Centre Chromosome 1
  Mapping Group. Further information can be found at
  http://www.sanger.ac.uk/HGP/Chr1
  438L4 is from the library RPC13 constructed at the Roswell Park
  Cancer Institute by the group of Pieter de Jong. For further
  details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.
FEATURES
    source
      1..87100
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="1"
        /map="p36.2-36.3"
        /clone="RP3-438L4"
        /clone_lib="RPC1-3"
        33..480
          /note="match: GSS B78720 clone 382B5"
          /note="L2 repeat: matches 2427..2508 of consensus"
          746..773
            /note="77 copies 4 mer caca 93% conserved"
            746..771
              /note="13 copies 2 mer ca 96% conserved"
              1352..2056
                /note="L1M5C repeat: matches 256..976 of consensus"
                2071..2227
                  /note="L1Mac repeat: matches 1557..1716 of consensus"
                  2228..2526
                    /note="AluX repeat: matches 2..300 of consensus"
                    2527..2709
                      /note="L1Mac repeat: matches 1716..1890 of consensus"
                      2719..2835
                        /note="L1M1 repeat: matches 6042..6157 of consensus"
                        2932..2984
                          /note="MIR repeat: matches 78..130 of consensus"
                          3286..3506
                            /note="AluSg/x repeat: matches 84..304 of consensus"
                            3517..3704
                              /note="LTR41 repeat: matches 597..772 of consensus"
                              4053..4193
                                /note="LTR41 repeat: matches 1..135 of consensus"
                                4911..5196
                                  /note="AluSg repeat: matches 1..284 of consensus"
                                  complement(<5103..>5396)
                                    /note="match: STS G07800"
                                    <5193..>5450
                                      /note="match: STS L30457"
                                      5197..5248
                                        /note="13 copies 4 mer taga 98% conserved"
                                        5197..5246
                                          /note="25 copies 2 mer ta 76% conserved"
                                          5426..5780
                                            /note="L2 repeat: matches 2335..2709 of consensus"
                                            6800..6807
                                              /note="Single clone region"
                                              6901..6954
                                                /note="27 copies 2 mer ta 80% conserved"
                                                6902..6957
                                                  /note="14 copies 4 mer atat 80% conserved"
                                                  7493..7746
                                                    /note="MER33 repeat: matches 5..258 of consensus"
                                                    7747..7889
                                                      /note="L1PA8 repeat: matches 6022..6163 of consensus"
                                                      7890..7953
                                                        /note="MER33 repeat: matches 258..322 of consensus"
                                                        9177..9674
                                                          /note="L1MB5 repeat: matches 5650..6176 of consensus"
                                                          9783..9866
                                                            /note="MIR repeat: matches 60..148 of consensus"
                                                            10330..10532
                                                              /note="MIR repeat: matches 13..224 of consensus"
                                                              11250..11290
                                                                /note="L2 repeat: matches 2665..2705 of consensus"
                                                                11947..12101
                                                                  /note="MIR repeat: matches 18..183 of consensus"
                                                                  12352..12693
                                                                    /note="L2 repeat: matches 2376..2707 of consensus"
                                                                    complement(<13712..>14212)
                                                                      /note="match: GSS AQ139995"
                                                                      15401..15769
                                                                        /note="MTII repeat: matches 57..410 of consensus"
                                                                        15905..16078
                                                                          /note="MERSA repeat: matches 1..183 of consensus"
                                                                          16546..16698
                                                                            /note="MIR repeat: matches 46..190 of consensus"
                                                                            17702..18055
                                                                              /note="L2 repeat: matches 2385..2749 of consensus"
                                                                              18228..18337
                                                                                /note="MIR repeat: matches 9..120 of consensus"
                                                                                complement(19963..20348)
                                                                                  /note="match: EST AA594177 clone IMAGE:1085527"
                                                                                  20428..20785
                                                                                    /note="L1MB3 repeat: matches 5813..6185 of consensus"
                                                                                    20855..20914
                                                                                      /note="30 copies 2 mer tg 80% conserved"
                                                                                      21820..22105
                                                                                        /note="AluX repeat: matches 1..286 of consensus"
                                                                                        22218..22368
                                                                                          /note="HAL1 repeat: matches 1427..1589 of consensus"
                                                                                          24465..24800
                                                                                            /note="2 copies 168 mer 79% conserved"
                                                                                            24561..24680
                                                                                              /note="60 copies 2 mer ca 60% conserved"
                                                                                              24727..24778
                                                                                                /note="26 copies 2 mer ca 71% conserved"
                                                                                                24779..24826
                                                                                                  /note="12 copies 4 mer acac 98% conserved"
```





TITLE	Genetic identification, sequence, and alternative splicing of the
JOURNAL	Caenorhabditis elegans alpha 2(IV) collagen gene
MEDLINE	J. Cell Biol. 123 (1), 255-264 (1993)
PUBMED	94012964
REFERENCE	7691828
AUTHORS	2 (bases 1 to 9541)
TITLE	Sibley, M.H.
JOURNAL	Direct Submission
FEATURES	Submitted (15-JUN-1993) Sibley M. H., Northwestern University
source	Medical School, Cell, Molecular and Structural Biology, 303 E.
	Chicago Ave., Chicago, Illinois, USA, 60611
	Location/Qualifiers
	1..9541
	/organism="Caenorhabditis elegans"
	/strain="N2 wild-type"
	/db_xref="taxon:6239"
5' UTR	187..695
intron	217..676
exon	/number=1
	677..718
	/number=2
mRNA	join(<596..718,936..990,1188..1250,1625..1724,2047..2235,
	2736..2911,2963..3040,3161..3268,3299..3409,3807..3923,
	3970..4134,4184..4455,4503..5364,5417..6089,6367..7181,
	7227..7458,7512..8303,8386..8849,8908..>9541)
	/product="a2(IV) collagen"
	/note="alternatively spliced transcript"
CDS	join(696..718,936..990,1188..1250,1625..1724,2047..2235,
	2736..2911,2963..3040,3161..3268,3807..3923,3970..4134,
	4184..4455,4503..5364,5417..6089,6367..7181,7227..7458,
	7512..8303,8386..8849,8908..9000)
	/note="alternatively spliced transcript"
	/codon_start=1
	/product="a2(IV) collagen"
	/protein_id="CAA80536.1"
	/db_xref="GI:953172"
	/db_xref="SWISS-PROT:P17140"
	/translation="MKQRAALGPVLRILAILALLAVSYVQSQATCRDCSNRGCFCVGEK
	GSMGAPGQPPGPTGIRGFPFEGLAGPKGLKGAQGGPPGVGKIDRGAVGVPFGP
	NDGNGRPEGPFGAGPAGWDCNGTGDGAPGIRGPPGMPGFPFGMCDLKGEPAL
	GYACAPGKGGMGMPGLPGPSRDGYGPEKGDGDTGNAGPRGPGGAGSGNPG
	IGSIGKGDGDLGSGVPPGPPGPREFTSGSIVGPRGNKEKDGKEPBGGRGIRP
	GNGLSQGLPGMKGKSLGSPAGPRGKEGRNAGPPGFKGDRGLDGLGIPGLPG
	KQKAGYVGRDRGNKSGPPGPGGGTNDGAPGGLGPRGNPPGPTDGYIPGAPG
	PAGTIGNTGGPGLPGYENGLPGPKGDKGDIIPGAPGVSGPSGIPGLPGKCEPY
	PTGTQSTIPLGPKRGKPLDAGPRKGENGLGVRGPPGDSNLNGLPCAGQRCAGP
	NGYDRGVNGLCAPCTKDRGCTCSACAPGTGKGLPGYSGQPPQDRGLPGHP
	GPVQADGDLGFPAGRPSPGPPQQDGFPLGPKQKEPTLTIRPPPPGYPGLKGN
	GFQPGQDGLPGSPGVPVPGAPGYPEKGDAGLPLGSKPGQDGLPLGPNKGEAG
	YGPQPGFFCAKGGDLPLGLPTGTLQGMGPEAPENQVNPAPPGQGLPLGPTGTG
	EGGYPRGPEVGPQFPGLPGMKGDSGLPGPPGLPHGVPDGKGVGVPGLGPIPGP
	KGDVGNPLGLNKGKGEVGVPGQPSGFPGLKGDAGLPLGTPGLGEGORFPG
	APGLKGDGLPLGLSQPGYGEKGDAGLPGVPRGSGFPFGQDGLPGVPMKGEDGL
	PLGPTVTLKGDLAGQSGAPLGAPYPMKNAGIPGVPGFKDGGDLPLGLPLGN
	GPKEGVPVPMGPTGPMKNGDLPLGRDGLSGVPMKGDGRGPNLPGKEGEPGA
	RDQKGDAGLPGQPLGPPQPSGLPGVFKGTGLPGVQPGQPGKEKGLGPIGKA
	GROCAPGSQDGLPGPMKGESYPGDGLPGRDGLPGVPGOKDGLGQSGQPLSG
	APGLDGQPGVPGIRGDKQGLPLGIPDRGMDGYPCQKGENGYPCQPLGLGKEGF
	ACTPFFGLKGSPPYPCQDGLPLGIPKGDGSGFPQPGQEGPLGSGEKGMLGMP
	QQPQSGTAGVPVPGAPGLQKQDGFPLGQKQGESLGLPGAPGLKGESMPGPGA
	KGDLGANGIPFKRGEDGLPGVPRDQGPPIGLKGEVGGAGLPGQPGFPPIGLKGE
	GLPGFCAKGAEPGTPGVGYAGEKGDGLPLGRDGLPGADGVPVPPGSPQON
	LVEPEKGLPLCAPGLRGEKGMPLDGPNDGPPGLGCRDNGDPCAPGLSGEK
	GMGLGFPGLDQPGPGAPGLPCAGCAAGPAYRDGFLVKHSQTTEVPRCPGQTK
	LWDGYSLIYIEGNEKSHNDLHAGSCLQRFSTMPFLFCDFNNVCNVRNEKSYWL
	TSEAIIPMNVNERIEPIYISRCACEAPANTIAVHSQTIQIPNCAGMSLWISYA
	MHTGAGEGGQSPSGSCLEDFRATPFTECNARGSCHYFANKFSFWLTITDNDSE
	FKVPESQTLKSGNLRVSRQVCVKSTDRGH"
	join(696..718,936..990,1188..1250,1625..1724,2047..2235,
	2736..2911,2963..3040,3299..3409,3807..3923,3970..4134,
	4184..4455,4503..5364,5417..6089,6367..7181,7227..7458,
	7512..8303,8386..8849,8908..9000)
	/note="alternatively spliced transcript"
CDS	join(696..718,936..990,1188..1250,1625..1724,2047..2235,
	2736..2911,2963..3040,3161..3268,3299..3409,3807..3923,
	3970..4134,4184..4455,4503..5364,5417..6089,6367..7181,
	7227..7458,7512..8303,8386..8849,8908..9000)
	/note="alternatively spliced transcript"
	/codon_start=1
	/product="a2(IV) collagen"
	/protein_id="CAA80537.1"
	/db_xref="GI:953173"
	/db_xref="SWISS-PROT:P17140"
	/translation="MKQRAALGPVLRILAILALLAVSYVQSQATCRDCSNRGCFCVGEK
	GSMGAPGQPPGPTGIRGFPFEGLAGPKGLKGAQGGPPGVGKIDRGAVGVPFGP
	NDGNGRPEGPFGAGPAGWDCNGTGDGAPGIRGPPGMPGFPFGMCDLKGEPAL
	GYACAPGKGGMGMPGLPGPSRDGYGPEKGDGDTGNAGPRGPGGAGSGNPG
	IGSIGKGDGDLGSGVPPGPPGPREFTSGSIVGPRGNKEKDGKEPBGGRGIRP
	GNGLSQGLPGMKGKSLGSPAGPRGKEGRNAGPPGFKGDRGLDGLGIPGLPG
	KQKAGYVGRDRGNKSGPPGPGGGTNDGAPGGLGPRGNPPGPTDGYIPGAPG
	PAGTIGNTGGPGLPGYENGLPGPKGDKGDIIPGAPGVSGPSGIPGLPGKCEPY
	PTGTQSTIPLGPKRGKPLDAGPRKGENGLGVRGPPGDSNLNGLPCAGQRCAGP
	NGYDRGVNGLCAPCTKDRGCTCSACAPGTGKGLPGYSGQPPQDRGLPGHP
	GPVQADGDLGFPAGRPSPGPPQQDGFPLGPKQKEPTLTIRPPPPGYPGLKGN
	GFQPGQDGLPGSPGVPVPGAPGYPEKGDAGLPLGSKPGQDGLPLGPNKGEAG
	YGPQPGFFCAKGGDLPLGLPTGTLQGMGPEAPENQVNPAPPGQGLPLGPTGTG
	EGGYPRGPEVGPQFPGLPGMKGDSGLPGPPGLPHGVPDGKGVGVPGLGPIPGP
	KGDVGNPLGLNKGKGEVGVPGQPSGFPGLKGDAGLPLGTPGLGEGORFPG
	APGLKGDGLPLGLSQPGYGEKGDAGLPGVPRGSGFPFGQDGLPGVPMKGEDGL
	PLGPTVTLKGDLAGQSGAPLGAPYPMKNAGIPGVPGFKDGGDLPLGLPLGN
	GPKEGVPVPMGPTGPMKNGDLPLGRDGLSGVPMKGDGRGPNLPGKEGEPGA
	RDQKGDAGLPGQPLGPPQPSGLPGVFKGTGLPGVQPGQPGKEKGLGPIGKA
	GROCAPGSQDGLPGPMKGESYPGDGLPGRDGLPGVPGOKDGLGQSGQPLSG
	APGLDGQPGVPGIRGDKQGLPLGIPDRGMDGYPCQKGENGYPCQPLGLGKEGF
	ACTPFFGLKGSPPYPCQDGLPLGIPKGDGSGFPQPGQEGPLGSGEKGMLGMP
	QQPQSGTAGVPVPGAPGLQKQDGFPLGQKQGESLGLPGAPGLKGESMPGPGA
	KGDLGANGIPFKRGEDGLPGVPRDQGPPIGLKGEVGGAGLPGQPGFPPIGLKGE
	GLPGFCAKGAEPGTPGVGYAGEKGDGLPLGRDGLPGADGVPVPPGSPQON
	LVEPEKGLPLCAPGLRGEKGMPLDGPNDGPPGLGCRDNGDPCAPGLSGEK
	GMGLGFPGLDQPGPGAPGLPCAGCAAGPAYRDGFLVKHSQTTEVPRCPGQTK
	LWDGYSLIYIEGNEKSHNDLHAGSCLQRFSTMPFLFCDFNNVCNVRNEKSYWL
	TSEAIIPMNVNERIEPIYISRCACEAPANTIAVHSQTIQIPNCAGMSLWISYA
	MHTGAGEGGQSPSGSCLEDFRATPFTECNARGSCHYFANKFSFWLTITDNDSE
	FKVPESQTLKSGNLRVSRQVCVKSTDRGH"
	join(696..718,936..990,1188..1250,1625..1724,2047..2235,
	2736..2911,2963..3040,3299..3409,3807..3923,3970..4134,
	4184..4455,4503..5364,5417..6089,6367..7181,7227..7458,
	7512..8303,8386..8849,8908..9000)
	/note="alternatively spliced transcript"
	/codon_start=1
	/product="a2(IV) collagen"
	/protein_id="CAA80537.1"
	/db_xref="GI:953173"
	/db_xref="SWISS-PROT:P17140"
	/translation="MKQRAALGPVLRILAILALLAVSYVQSQATCRDCSNRGCFCVGEK
	GSMGAPGQPPGPTGIRGFPFEGLAGPKGLKGAQGGPPGVGKIDRGAVGVPFGP
	NDGNGRPEGPFGAGPAGWDCNGTGDGAPGIRGPPGMPGFPFGMCDLKGEPAL
	GYACAPGKGGMGMPGLPGPSRDGYGPEKGDGDTGNAGPRGPGGAGSGNPG
	IGSIGKGDGDLGSGVPPGPPGPREFTSGSIVGPRGNKEKDGKEPBGGRGIRP
	GNGLSQGLPGMKGKSLGSPAGPRGKEGRNAGPPGFKGDRGLDGLGIPGLPG
	KQKAGYVGRDRGNKSGPPGPGGGTNDGAPGGLGPRGNPPGPTDGYIPGAPG
	PAGTIGNTGGPGLPGYENGLPGPKGDKGDIIPGAPGVSGPSGIPGLPGKCEPY
	PTGTQSTIPLGPKRGKPLDAGPRKGENGLGVRGPPGDSNLNGLPCAGQRCAGP
	NGYDRGVNGLCAPCTKDRGCTCSACAPGTGKGLPGYSGQPPQDRGLPGHP
	GPVQADGDLGFPAGRPSPGPPQQDGFPLGPKQKEPTLTIRPPPPGYPGLKGN
	GFQPGQDGLPGSPGVPVPGAPGYPEKGDAGLPLGSKPGQDGLPLGPNKGEAG
	YGPQPGFFCAKGGDLPLGLPTGTLQGMGPEAPENQVNPAPPGQGLPLGPTGTG
	EGGYPRGPEVGPQFPGLPGMKGDSGLPGPPGLPHGVPDGKGVGVPGLGPIPGP
	KGDVGNPLGLNKGKGEVGVPGQPSGFPGLKGDAGLPLGTPGLGEGORFPG
	APGLKGDGLPLGLSQPGYGEKGDAGLPGVPRGSGFPFGQDGLPGVPMKGEDGL
	PLGPTVTLKGDLAGQSGAPLGAPYPMKNAGIPGVPGFKDGGDLPLGLPLGN
	GPKEGVPVPMGPTGPMKNGDLPLGRDGLSGVPMKGDGRGPNLPGKEGEPGA
	RDQKGDAGLPGQPLGPPQPSGLPGVFKGTGLPGVQPGQPGKEKGLGPIGKA
	GROCAPGSQDGLPGPMKGESYPGDGLPGRDGLPGVPGOKDGLGQSGQPLSG
	APGLDGQPGVPGIRGDKQGLPLGIPDRGMDGYPCQKGENGYPCQPLGLGKEGF
	ACTPFFGLKGSPPYPCQDGLPLGIPKGDGSGFPQPGQEGPLGSGEKGMLGMP
	QQPQSGTAGVPVPGAPGLQKQDGFPLGQKQGESLGLPGAPGLKGESMPGPGA
	KGDLGANGIPFKRGEDGLPGVPRDQGPPIGLKGEVGGAGLPGQPGFPPIGLKGE
	GLPGFCAKGAEPGTPGVGYAGEKGDGLPLGRDGLPGADGVPVPPGSPQON
	LVEPEKGLPLCAPGLRGEKGMPLDGPNDGPPGLGCRDNGDPCAPGLSGEK
	GMGLGFPGLDQPGPGAPGLPCAGCAAGPAYRDGFLVKHSQTTEVPRCPGQTK
	LWDGYSLIYIEGNEKSHNDLHAGSCLQRFSTMPFLFCDFNNVCNVRNEKSYWL
	TSEAIIPMNVNERIEPIYISRCACEAPANTIAVHSQTIQIPNCAGMSLWISYA
	MHTGAGEGGQSPSGSCLEDFRATPFTECNARGSCHYFANKFSFWLTITDNDSE
	FKVPESQTLKSGNLRVSRQVCVKSTDRGH"
	join(696..718,936..990,1188..1250,1625..1724,2047..2235,
	2736..2911,2963..3040,3299..3409,3807..3923,3970..4134,
	4184..4455,4503..5364,5417..6089,6367..7181,7227..7458,
	7512..8303,8386..8849,8908..9000)
	/note="alternatively spliced transcript"
	/codon_start=1
	/product="a2(IV) collagen"
	/protein_id="CAA80537.1"
	/db_xref="GI:953173"
	/db_xref="SWISS-PROT:P17140"
	/translation="MKQRAALGPVLRILAILALLAVSYVQSQATCRDCSNRGCFCVGEK
	GSMGAPGQPPGPTGIRGFPFEGLAGPKGLKGAQGGPPGVGKIDRGAVGVPFGP
	NDGNGRPEGPFGAGPAGWDCNGTGDGAPGIRGPPGMPGFPFGMCDLKGEPAL
	GYACAPGKGGMGMPGLPGPSRDGYGPEKGDGDTGNAGPRGPGGAGSGNPG
	IGSIGKGDGDLGSGVPPGPPGPREFTSGSIVGPRGNKEKDGKEPBGGRGIRP
	GNGLSQGLPGMKGKSLGSPAGPRGKEGRNAGPPGFKGDRGLDGLGIPGLPG
	KQKAGYVGRDRGNKSGPPGPGGGTNDGAPGGLGPRGNPPGPTDGYIPGAPG
	PAGTIGNTGGPGLPGYENGLPGPKGDKGDIIPGAPGVSGPSGIPGLPGKCEPY
	PTGTQSTIPLGPKRGKPLDAGPRKGENGLGVRGPPGDSNLNGLPCAGQRCAGP
	NGYDRGVNGLCAPCTKDRGCTCSACAPGTGKGLPGYSGQPPQDRGLPGHP
	GPVQADGDLGFPAGRPSPGPPQQDGFPLGPKQKEPTLTIRPPPPGYPGLKGN
	GFQPGQDGLPGSPGVPVPGAPGYPEKGDAGLPLGSKPGQDGLPLGPNKGEAG
	YGPQPGFFCAKGGDLPLGLPTGTLQGMGPEAPENQVNPAPPGQGLPLGPTGTG
	EGGYPRGPEVGPQFPGLPGMKGDSGLPGPPGLPHGVPDGKGVGVPGLGPIPGP
	KGDVGNPLGLNKGKGEVGVPGQPSGFPGLKGDAGLPLGTPGLGEGORFPG
	APGLKGDGLPLGLSQPGYGEKGDAGLPGVPRGSGFPFGQDGLPGVPMKGEDGL
	PLGPTVTLKGDLAGQSGAPLGAPYPMKNAGIPGVPGFKDGGDLPLGLPLGN
	GPKEGVPVPMGPTGPMKNGDLPLGRDGLSGVPMKGDGRGPNLPGKEGEPGA
	RDQKGDAGLPGQPLGPPQPSGLPGVFKGTGLPGVQPGQPGKEKGLGPIGKA
	GROCAPGSQDGLPGPMKGESYPGDGLPGRDGLPGVPGOKDGLGQSGQPLSG
	APGLDGQPGVPGIRGDKQGLPLGIPDRGMDGYPCQKGENGYPCQPLGLGKEGF
	ACTPFFGLKGSPPYPCQDGLPLGIPKGDGSGFPQPGQEGPLGSGEKGMLGMP
	QQPQSGTAGVPVPGAPGLQKQDGFPLGQKQGESLGLPGAPGLKGESMPGPGA
	KGDLGANGIPFKRGEDGLPGVPRDQGPPIGLKGEVGGAGLPGQPGFPPIGLKGE
	GLPGFCAKGAEPGTPGVGYAGEKGDGLPLGRDGLPGADGVPVPPGSPQON
	LVEPEKGLPLCAPGLRGEKGMPLDGPNDGPPGLGCRDNGDPCAPGLSGEK
	GMGLGFPGLDQPGPGAPGLPCAGCAAGPAYRDGFLVKHSQTTEVPRCPGQTK
	LWDGYSLIYIEGNEKSHNDLHAGSCLQRFSTMPFLFCDFNNVCNVRNEKSYWL
	TSEAIIPMNVNERIEPIYISRCACEAPANTIAVHSQTIQIPNCAGMSLWISYA
	MHTGAGEGGQSPSGSCLEDFRATPFTECNARGSCHYFANKFSFWLTITDNDSE
	FKVPESQTLKSGNLRVSRQVCVKSTDRGH"
	join(696..718,936..990,1188..1250,1625..1724,2047..2235,
	2736..2911,2963..3040,3299..3409,3807..3923,3970..4134,
	4184..4455,4503..5364,5417..6089,6367..7181,7227..7458,
	7512..8303,8386..8849,8908..9000)
	/note="alternatively spliced transcript"
	/codon_start=1
	/product="a2(IV) collagen"
	/protein_id="CAA80537.1"
	/db_xref="GI:953173"
	/db_xref="SWISS-PROT:P17140"
	/translation="MKQRAALGPVLRILAILALLAVSYVQSQATCRDCSNRGCFCVGEK
	GSMGAPGQPPGPTGIRGFPFEGLAGPKGLKGAQGGPPGVGKIDRGAVGVPFGP
	NDGNGRPEGPFGAGPAGWDCNGTGDGAPGIRGPPGMPGFPFGMCDLKGEPAL
	GYACAPGKGGMGMPGLPGPSRDGYGPEKGDGDTGNAGPRGPGGAGSGNPG
	IGSIGKGDGDLGSGVPPGPPGPREFTSGSIVGPRGNKEKDGKEPBGGRGIRP
	GNGLSQGLPGMKGKSLGSPAGPRGKEGRNAGPPGFKGDRGLDGLGIPGLPG
	KQKAGYVGRDRGNKSGPPGPGGGTNDGAPGGLGPRGNPPGPTDGYIPGAPG
	PAGTIGNTGGPGLPGYENGLPGPKGDKGDIIPGAPGVSGPSGIPGLPGKCEPY
	PTGTQSTIPLGPKRGKPLDAGPRKGENGLGVRGPPGDSNLNGLPCAGQRCAGP
	NGYDRGVNGLCAPCTKDRGCTCSACAPGTGKGLPGYSGQPPQDRGLPGHP
	GPVQADGDLGFPAGRPSPGPPQQDGFPLGPKQKEPTLTIRPPPPGYPGLKGN
	GFQPGQDGLPGSPGVPVPGAPGYPEKGDAGLPLGSKPGQDGLPLGPNKGEAG
	YGPQPGFFCAKGGDLPLGLPTGTLQGMGPEAPENQVNPAPPGQGLPLGPTGTG
	EGGYPRGPEVGPQFPGLPGMKGDSGLPGPPGLPHGVPDGKGVGVPGLGPIPGP
	KGDVGNPLGLNKGKGEVGVPGQPSGFPGLKGDAGLPLGTPGLGEGORFPG
	APGLKGDGLPLGLSQPGYGEKGDAGLPGVPRGSGFPFGQDGLPGVPMKGEDGL
	PLGPTVTLKGDLAGQSGAPLGAPYPMKNAGIPGVPGFKDGGDLPLGLPLGN
	GPKEGVPVPMGPTGPMKNGDLPLGRDGLSGVPMKGDGRGPNLPGKEGEPGA
	RDQKGDAGLPGQPLGPPQPSGLPGVFKGTGLPGVQPGQPGKEKGLGPIGKA
	GROCAPGSQDGLPGPMKGESYPGDGLPGRDGLPGVPGOKDGLGQSGQPLSG
	APGLDGQPGVPGIRGDKQGLPLGIPDRGMDGYPCQKGENGYPCQPLGLGKEGF
	ACTPFFGLKGSPPYPCQDGLPLGIPKGDGSGFPQPGQEGPLGSGEKGMLGMP
	QQPQSGTAGVPVPGAPGLQKQDGFPLGQKQGESLGLPGAPGLKGESMPGPGA
	KGDLGANGIPFKRGEDGLPGVPRDQGPPIGLKGEVGGAGLPGQPGFPPIGLKGE
	GLPGFCAKGAEPGTPGVGYAGEKGDGLPLGRDGLPGADGVPVPPGSPQON
	LVEPEKGLPLCAPGLRGEKGMPLDGPNDGPPGLGCRDNGDPCAPGLSGEK
	GMGLGFPGLDQPGPGAPGLPCAGCAAGPAYRDGFLVKHSQTTEVPRCPGQTK
	LWDGYSLIYIEGNEKSHNDLHAGSCLQRFSTMPFLFCDFNNVCNVRNEKSYWL
	TSEAIIPMNVNERIEPIYISRCACEAPANTIAVHSQTIQIPNCAGMSLWISYA
	MHTGAGEGGQSPSGSCLEDFRATPFTECNARGSCHYFANKFSFWLTITDNDSE
	FKVPESQTLKSGNLRVSRQVCVKSTDRGH"
	join(696..718,936..990,1188..1250,1625..1724,2047..2235,
	2736..2911,2963..3040,3299..3409,3807..3923,3970..4134,
	4184..4455,4503..5364,5417..6089,6367..7181,7227..7458,
	7512..8303,8386..8849,8908..9000)
	/note="alternatively spliced transcript"
	/codon_start=1
	/product="a2(IV) collagen"
	/protein_id="CAA80537.1"
	/db_xref="GI:953173"
	/db_xref="SWISS-PROT:P17140"
	/translation="MKQRAALGPVLRILAILALLAVSYVQSQATCRDCSNRGCFCVGEK
	GSMGAPGQPPGPTGIRGFPFEGLAGPKGLKGAQGGPPGVGKIDRGAVGVPFGP
	NDGNGRPEGPFGAGPAGWDCNGTGDGAPGIRGPPGMPGFPFGMCDLKGEPAL
	GYACAPGKGGMGMPGLPGPSRDGYGPEKGDGDTGNAGPRGPGGAGSGNPG
	IGSIGKGDGDLGSGVPPGPPGPREFTSGSIVGPRGNKEKDGKEPBGGRGIRP
	GNGLSQGLPGMKGKSLGSPAGPRGKEGRNAGPPGFKGDRGLDGLGIPGLPG
	KQKAGYVGRDRGNKSGPPGPGGGTNDGAPGGLGPRGNPPGPTDGYIPGAPG
	PAGTIGNTGGPGLPGYENGLPGPKGDKGDIIPGAPGVSGPSGIPGLPGKCEPY
	PTGTQSTIPLGPKRGKPLDAGPRKGENGLGVRGPPGDSNLNGLPCAGQRCAGP
	NGYDRGVNGLCAPCTKDRGCTCSACAPGTGKGLPGYSGQPPQDRGLPGHP
	GPVQADGDLGFPAGRPSPGPPQQDGFPLGPKQKEPTLTIRPPPPGYPGLKGN
	GFQPGQDGLPGSPGVPVPGAPGYPEKGDAGLPLGSKPGQDGLPLGPNKGEAG
	YGPQPGFFCAKGGDLPLGLPTGTLQGMGPEAPENQVNPAPPGQGLPLGPTGTG
	EGGYPRGPEVGPQFPGLPGMKGDSGLPGPPGLPHGVPDGKGVGVPGLGPIPGP
	KGDVGNPLGLNKGKGEVGVPGQPSGFPGLKGDAGLPLGTPGLGEGORFPG
	APGLKGDGLPLGLSQPGYGEKGDAGLPGVPRGSGFPFGQDGLPGVPMKGEDGL
	PLGPTVTLKGDLAGQSGAPLGAPYPMKNAGIPGVPGFKDGGDLPLGLPLGN
	GPKEGVPVPMGPTGPMKNGDLPLGRDGLSGVPMKGDGRGPNLPGKEGEPGA
	RDQKGDAGLPGQPLGPPQPSGLPGVFKGTGLPGVQPGQPGKEKGLGPIGKA
	GROCAPGSQDGLPGPMKGESYPGDGLPGRDGLPGVPGOKDGLGQSGQPLSG
	APGLDGQPGVPGIRGDKQGLPLGIPDRGMDGYPCQKGENGYPCQPLGLGKEGF
	ACTPFFGLKGSPPYPCQDGLPLGIPKGDGSGFPQPGQEGPLGSGEKGMLGMP
	QQPQSGTAGVPVPGAPGLQKQDGFPLGQKQGESLGLPGAPGLKGESMPGPGA
	KGDLGANGIPFKRGEDGLPGVPRDQGPPIGLKGEVGGAGLPGQPGFPPIGLKGE
	GLPGFCAKGAEPGTPGVGYAGEKGDGLPLGRDGLPGADGVPVPPGSPQON
	LVEPEKGLPLCAPGLRGEKGMPLDGPNDGPPGLGCRDNGDPCAPGLSGEK
	GMGLGFPGLDQPGPGAPGLPCAGCAAGPAYRDGFLVKHSQTTEVPRCPGQTK
	LWDGYSLIYIEGNEKSHNDLHAGSCLQRFSTMPFLFCDFNNVCNVRNEKSYWL
	TSEAIIPMNVNERIEPIYISRCACEAPANTIAVHSQTIQIPNCAGMSLWISYA
	MHTGAGEGGQSPSGSCLEDFRATPFTECNARGSCHYFANKFSFWLTITDNDSE
	FKVPESQTLKSGNLRVSRQVCVKSTDRGH"
	join(696..718,936..990,1188..1250,1625..1724,2047..2235,
	2736..2911,2963..3040,3299..3409,3807..3923,3970..4134,
	4184..4455,4503..5364,5417..6089,6367..7181,7227..7458,
	7512..8303,8386..8849,8908..9000)
	/note="alternatively spliced transcript"
	/codon_start=1
	/product="a2(IV) collagen"
	/protein_id="CAA80537.1"
	/db_xref="GI:953173"
	/db_xref="SWISS-PROT:P17140"
	/translation="MKQRAALGPVLRILAILALLAVSYVQSQATCRDCSNRGCFCVGEK
	GSMGAPGQPPGPTGIRGFPFEGLAGPKGLKGAQGGPPGVGKIDRGAVGVPFGP
	NDGNGRPEGPFGAGPAGWDCNGTGDGAPGIRGPPGMPGFPFGMCDLKGEPAL
	GYACAPGKGGMGMPGLPGPSRDGYGPEKGDGDTGNAGPRGPGGAGSGNPG
	IGSIGKGDGDLGSGVPPGPPGPREFTSGSIVGPRGNKEKDGKEPBGGRGIRP
	GNGLSQGLPGMKGKSLGSPAGPRGKEGRNAGPPGFKGDRGLDGLGIPGLPG
	KQKAGYVGRDRGNKSGPPGPGGGTNDGAPGGLGPRGNPPGPTDGYIPGAPG
	PAGTIGNTGGPGLPGYENGLPGPKGDKGDIIPGAPGVSGPSGIPGLPGKCEPY
	PTGTQSTIPLGPKRGKPLDAGPRKGENGLGVRGPPGDSNLNGLPCAGQRCAGP
	NGYDRGVNGLCAPCTKDRGCTCSACAPGTGKGLPGYSGQPPQDRGLPGHP
	GPVQADGDLGFPAGRPSPGPPQQDGFPLGPKQKEPTLTIRPPPPGYPGLKGN
	GFQPGQDGLPGSPGVPVPGAPGYPEKGDAGLPLGSKPGQDGLPLGPNKGEAG
	YGPQPGFFCAKGGDLPLGLPTGTLQGMGPEAPENQVNPAPPGQGLPLGPTGTG
	EGGYPRGPEVGPQFPGLPGMKGDSGLPGPPGLPHGVPDGKGVGVPGLGPIPGP
	KGDVGNPLGLNKGKGEVGVPGQPSGFPGLKGDAGLPLGTPGLGEGORFPG





```
Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Guan,X., Gupta,J.,
Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
Marquies,E.H., Mastello,C., Maskeri,B., Mastrian,S.D.,
McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
Thomas,P.J., Touchman,J.W., Tsurgon,C., Vogt,J.L., Walker,M.A.,
Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
unpublished
2 (bases 1 to 123789)
Green,E.D.
Direct Submission
Submitted (14-FEB-2002) NIH Intramural Sequencing Center, 8717
Grovesmont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 123789)
Green,E.D.
Direct Submission
Submitted (18-JUL-2002) NIH Intramural Sequencing Center, 8717
Grovesmont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov
----- Project Information
Center project name: cvk
Center clone name: 043116
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 123784 bases at least Q40
Consensus quality: 123789 bases at least Q30
Consensus quality: 123789 bases at least Q20
Insert size: 131000; agarose-fp
Quality coverage: 10.94x in Q20 bases; agarose-fp
Quality coverage: 11.58x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1 123789: contig of 123789 bp in length.
FEATURES             source
    1..123789
        /location=Qualifiers
        /organism="Tetraodon nigroviridis"
        /db_xref="taxon:99883"
        /clone="GSTNA-43116"
        /clone_lib="GSTNA"
        /note="Genoscope designation: C0AA043116"
misc_feature         1..123789
        /note="assembly_fragment
        clone_end:SP6
        vector_side:right"
BASE COUNT          35968 a 26646 c 26925 g 34250 t
ORIGIN
Query Match          24.0%; Score 37.2; DB 2; Length 123789;
Best Local Similarity 68.9%; Pred. No. 1.8;
Matches 51; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
QY 43 GTAGAGATTGCTGGACCTCCAGCACACCCAGCCGCCAGAGAAAGTGGGGCTCTCGT 102
||||| ||||| ||||| ||| ||| |||| ||||| ||||| ||||| ||||| |||||
Db 355 GTGGGATTCCTGGACCTCAGGTCATACAGGCCCTGAGGAGCCATGGGGCCACCCGA 296
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 103 GCACCAAGTTTACC 116
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

---

```
Db 295 GCACCTGGTTTACC 282
RESULT 10
MMCOLA4
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES             source
    1..3612
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /clone="pCIV-1-PE12, pCIV-1-PE16"
        /clone_lib="PYS cdna"
        /note="uc was cc in [1]"
        /citation=[1]
old_sequence        151..>3612
        /note="collagen alpha chain precursor (AA -27 to 1127)"
        /codon_start=1
        /protein_id="CAA29946.1"
        /db_xref="GI:50502"
        /db_xref="MGD:MGI:88454"
        /translation="MGPRLSVWLLLLFAALLLHEERSRAPAKGDCGGSCGKDCDHG
        KGKGERGLPGLVGIPGMOGPEGHCPGCKGKDGACGPGI.PGKTGTRGPPCAAGYP
        GNPLGPIPGQDPPGPGIPGICNGTKGERGLGPPGLPGFSGNPGPGLPMKMGDGP
        EILHVPGLTLKRGFPGIPGMPGLPLGLQGVPPGFTGPGPPGPPGPGKEG
        QMGSSFGPKGDKGQGVGPGVQAQKESKGFAPTKGKQGEKGFPGVPGYGE
        IGTPLGKGRDGYPCAPLGRGEPGKPGFTPLGPPGFTPGQAAGAPGPPGGE
        KGDCFGPVSPGSRDAGPPGPGPGPGPGPGHTNGIVECQPGPGDQGPPTGQ
        PGLTGEVQKQKGSCLACDTGLRGLGPPGLQGPGEIGFPGHPCAKGDRGLPGRDGL
        EGLPGQSGPLIGTGAKGEGEIFFDMRLKDKDGPFGPQPGMPGAGTGGRRDGH
        PGLPGKSPGSGISGLKGERGPGGVPFGSRGIDGPPGPGVGI.PVGEKQAGFPFG
        GPGSPGLPGKGAAGVPLPGPGAAGLPGSPGPPGPGDRGFTGPGTGPPIGPKG
        AVGPGIGFPGPLPGKGVGPGVGGGLTGPGLGIRGDPGPGVQGPAGPPGV
        LKGPGLPIPTGEGKSGISGPGVGGGLTGPGLGIRGDPGPGVQGPAGPPGV
        PGIPPPGAWPPGGGPPGKSGSGPPGPKGEGFPFGPLDMPGKDGKSQGLGLTGQ
        SGLPGLPQQGTPPGVPFPGSKGVMGTGPGQSPGPGTGPGLPGEKDGSLPGSS
        GPRGDPGKDGKDVGLPGMPGSMHVDMSMKQKGDQKGGKGLGTPGDKSGRDPG
        TPVPGKDHAGHPGQPGPKGDPGLSGTPGSPGLPGKSGVGMGLPGSGKSGVPGI
        PSCQVPGSPGKGAKEKQSGSLPGIGIPGRPGKDGQGLAGFPGSPGKSGKSGAG
        TPGMGSPGSPGSPGNHPSGPLGPKGDKGLPGLDGVPGVKGAGLPGTPTGTP
        AqKGEPC
        151..231
        232..882
sig_peptide
mat_peptide
```

```
mat_peptide      /product="7S protein (AA 1-217)"
232..276
old_sequence     /product="N-telopeptide (AA 1-15)"
181
/note="c was g in [1]"
/citation=[1]
misc_feature     883..3612
/note="helical region (AA 218-1127)"
misc_feature     883..921
/note="NC2 domain (AA 218-230)"
misc_feature     3475..3555
/note="exon 4 region"
BASE COUNT      819 a 1011 c 1212 g 570 t
ORIGIN
Query Match      23.7%; Score 36.8; DB 10; Length 3612;
Best Local Similarity 61.5%; Pred. No. 1.7;
Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
OY 53 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGCCCTCCCTGGTGACACAGGTT 112
Db 2711 CAGGACAGTTCAGGCTCCCTGGCTTCTTCTGGACAGCAGGAGCACCTGGAGTTCACGGGT 2770
OY 113 TACCACAATATACAGGAGAAATAGTGAATGACAA 148
Db 2771 TCCAGGTTCTAAAGGTGAATGGTGTCTATGGGAA 2806
RESULT 11
AX306181 LOCUS AX306181 6512 bp DNA linear PAT 11-DEC-2001
DEFINITION Sequence 932 from Patent WO0188188.
ACCESSION AX306181
VERSION AX306181.1 GI:17645461
KEYWORDS house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Ishikawa,K., Asai,S., Takahashi,Y., Nagata,T. and Ishii,Y.
TITLE Method for examining ischemic conditions
JOURNAL Patent: WO 018188-A 932 22-NOV-2001;
School Juridical Person Nihon University (JP)
FEATURES
Location/Qualifiers
source
1..6512
/organism="Mus musculus"
/db_xref="taxon:10090"
BASE COUNT 1553 a 1798 c 1950 g 1211 t
ORIGIN
Query Match      23.7%; Score 36.8; DB 6; Length 6512;
Best Local Similarity 61.5%; Pred. No. 1.8;
Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
OY 53 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGCCCTCCCTGGTGACACAGGTT 112
Db 2692 CAGGACAGTTCAGGCTCCCTGGCTTCTTCTGGACAGCAGGAGCACCTGGAGTTCACGGGT 2751
OY 113 TACCACAATATACAGGAGAAATAGTGAATGACAA 148
Db 2752 TCCAGGTTCTAAAGGTGAATGGTGTCTATGGGAA 2787
RESULT 12
MUSCOL1A4A LOCUS MUSCOL1A4A 6512 bp mRNA linear ROD 06-OCT-1994
DEFINITION Mus musculus alpha-1 type IV collagen (Col4a-1) mRNA, complete cds.
ACCESSION J04694
VERSION J04694.1 GI:556296
KEYWORDS alpha-1 type IV collagen.
SOURCE Mus musculus 13.5 day old embryo parietal endoderm cdna to mRNA.
ORGANISM Mus musculus
```

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## COMMENT

## FEATURES

## source

## gene

## CDS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 6512)

Muthukumar, G., Blumberg, B. and Kurkinen, M.

The complete primary structure for the alpha 1-chain of mouse collagen IV. Differential evolution of collagen IV domains

J. Biol. Chem. 264 (11), 6310-6317 (1989)

89197932

2703490

On Oct 8, 1994 this sequence version replaced gi:340547.

Location/Qualifiers

1..6512

/organism="Mus musculus"

/db\_xref="taxon:10090"

/tissue\_type="parietal endoderm"

/dev\_stage="13.5 day old embryo"

1..6512

/gene="Col4a-1"

132..5141

/gene="Col4a-1"

/codon\_start=1

/product="alpha-1 type IV collagen"

/protein\_id="AA50292.1"

/db\_xref="GI:556297"

/translation="MGPRLSVLLLLFAALLLHEERSRAAKDCGGSGCKKCDCHGV

KNGKGERGLPGLQVIGFPMGMPGEGPHGPGKQGDAGEPLGCTKGTGRPPGGAAGYP

GNGLPLGIPQDGPDPGPPGIPGCGTKGERGLPGLPGFSGNPGPGLPMGMDPG

EILGHVPTLLKGERFPPGIPGMPSGLPLOGVPGPFTGPPGPPGPPGPPGPPG

OMGSSFGQKDGKEQGVSGPVGAQVKEGDFAPTEKGGKGGPPGPPGPPGPPG

KGPPGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGG

IGTMLPEKGDGYPGAPLGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG

KDQGLPGSLPGRSDGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG

PGLTGVGQKQKESCLACTGELRGLPQGGPPGPPGPPGPPGPPGPPGPPGPPG

EGLPGQSGPLIGQCAKGEPEIFDMLKGDGDPGPPGPPGPPGPPGPPGPPGPPG

PGLPGKSGPSGLKGERGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG

GPSPGLPFPKGEAGKVYPLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLP

AVQPGIGFPLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLP

LKQGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGL

PGTGPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLT

SGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLT

BRGDPLGKDGKDVGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLP

TPGVPKDQAGHPQPGKDPGLSGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGL

PGSGVPSGPKGAKGKGGSLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLP

TPQMPSGPRSGSNLHPSGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLP

AGKGEPPGSDGIPGSAGKEGQGVPGPPGPPGPPGPPGPPGPPGPPGPPGPPG

VKGEQFMGPQGGQGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLP

GPKDKNGMPCAGVPKPGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGG

LPGLQVKGQDQGVPGPKLQGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLP

PKQQGVTVSGVLPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP

PGTSPVDHGLFVTRHSQTDPLCPGPKILYHGYSLLYVQGNRAHGDGLGTAGSCL

RKFTMPFLFCNINNVNCFASRNDYSYWLSTPEPMPMSAPISGDNTRPFTSRCAVCE

APAMVAVHSOTIQIPCPNGWSSLIWYGSFVMTSAGASGQALASPSCLEFRS

APTECHGRGTCNYANAYSWLATIERSEMKKPTSTLKAGELRTHVSRQVCMRR

"

132..212

/gene="Col4a-1"

213..5138

/gene="Col4a-1"

/product="alpha-1 type IV collagen"

5406..5411

/gene="Col4a-1"

/note="putative"

5732..5737

/gene="Col4a-1"

/note="putative"

5761..5766

/gene="Col4a-1"

/note="putative"

5765..5770

/gene="Col4a-1"

/note="putative"

5777..5782

/gene="Col4a-1"

sig\_peptide

mat\_peptide

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal



	/product="mini-collagen" /protein_id="CAA43379.1" /db_xref="GI:9447" /db_xref="SPTREMBL:Q00484" /translation="NAMRLVLAKTLVGAATQAOKTLEHMLKRDANPCSGYCSPVCAPA CAPVCCYPP GPPGPGAPGGGLPQQAPPPPPPPPPPPPPPPPPPPPPPVCAOCVTPCPOYCPAKRK" 14..73 /gene="N-COL 1" 74..460 /gene="N-COL 1" /product="mini-collagen" 167..235 /note="proline repeat" 248..373 /gene="N-COL 1" /note="collagen-like region"	BASE COUNT 218 a 197 c 123 g 156 t	TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL
	Query Match 23.6%; Score 36.6; DB 3; Length 694; Best Local Similarity 62.6%; Pred. No. 1-7;		
Matches	57; Conservative 0; Mismatches 34; Indels 0; Gaps 0;		
QY 51	TGCTGCACCTCCAGCACACCCAGGCCCGCAGAAGTAGTGCGGTGCACCAGG 110   		REFERENCE AUTHORS TITLE JOURNAL
Dd 253	TCCAGGACCTCCAGGACCTCAGGACGACGACGACGACGACGACGACGACGACG 312 		
QY 111	TTTACCACATATACAGGAGAAATAAGTGAA 141 		COMMENT
Dd 313	ACTTCAGGACCTCCAGGACCCACGAGCA 343 		
RESULT 15			
LOCUS AC096135/c			
DEFINITION Rattus norvegicus clone CH230-11E5, *** SEQUENCING IN PROGRESS ***,	95492 bp DNA linear HTG 11-JUL-2002		
LOCUS AC096135	46 unordered pieces.		
ACCESSION AC096135.3 GI:21723277			
VERSION HTG: HTGS_PHASE1.			
KEYWORDS Norway rat.			
SOURCE Rattus norvegicus			
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
REFERENCE 1 (bases 1 to 95492)			
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amarutunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbarta,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Joliviet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kufeshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Lounseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,			
	Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oraqunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojebokan,I., Rolfe,M., Ruiz,S., Savory,G., Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Telfrod,B., Thomas,N., Thomas,S., Tansey,J., Taylor,C., Taylor,T., Vera,V., Villalon,D., Vinson,R., Wang,Q., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wluczka,R., Woodden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R. Direct Submission Unpublished 2 {bases 1 to 95492} Worley,K.C. Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 3 {bases 1 to 95492} Worley,K.C. Submitted (11-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Jul 10, 2002 this sequence version replaced gl:17943819. ----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu ----- Project Information Center project name: GENI Center clone name: CH230-11E5 ----- Summary Statistics Sequencing vector: Plasmid; Chemistry: Dye-terminator Big Dye; 100% of reads Assembly program: Phrap; version 0.990329 Consensus quality: 36046 bases at least Q40 Consensus quality: 42440 bases at least Q30 Consensus quality: 46490 bases at least Q20 ----- ** NOTE: Estimated insert size may differ from sequence length ** (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html). ** NOTE: This is a 'working draft' sequence. It currently consists of 46 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. ** This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. * 1 1132: contig of 1132 bp in length * 1133 1232: gap of unknown length * 1233 2500: contig of 1268 bp in length * 2501 2600: gap of unknown length * 2601 4032: contig of 1432 bp in length * 4033 4132: gap of unknown length * 4133 5210: contig of 1078 bp in length * 5211 5310: gap of unknown length * 5311 6685: contig of 1375 bp in length * 6686 6785: gap of unknown length * 6786 8171: contig of 1386 bp in length * 8172 8271: gap of unknown length * 8272 9447: contig of 1176 bp in length * 9448 9548: gap of unknown length * 9548 11015: contig of 1468 bp in length * 11016 11115: gap of unknown length * 11116 12931: contig of 1816 bp in length * 12932 13031: gap of unknown length		

\* 13032 14745: contig of 1714 bp in length  
\* 14746 14845: gap of unknown length  
\* 14846 16501: contig of 1656 bp in length  
\* 16502 16601: gap of unknown length  
\* 16602 17952: contig of 1351 bp in length  
\* 17953 18052: gap of unknown length  
\* 18053 19788: contig of 1736 bp in length  
\* 19789 19888: gap of unknown length  
\* 19889 21580: contig of 1692 bp in length  
\* 21581 21680: gap of unknown length  
\* 21681 22901: contig of 1221 bp in length  
\* 22902 23001: gap of unknown length  
\* 23002 24262: contig of 1261 bp in length  
\* 24263 24362: gap of unknown length  
\* 24363 25367: contig of 1005 bp in length  
\* 25368 25467: gap of unknown length  
\* 25468 25915: contig of 1448 bp in length  
\* 25916 27015: gap of unknown length  
\* 27016 28543: contig of 1528 bp in length  
\* 28544 28643: gap of unknown length  
\* 28644 30278: contig of 1635 bp in length  
\* 30279 30378: gap of unknown length  
\* 30379 31382: contig of 1004 bp in length  
\* 31383 31482: gap of unknown length  
\* 31483 33124: contig of 1642 bp in length  
\* 33125 33224: gap of unknown length  
\* 33225 34854: contig of 1630 bp in length  
\* 34855 34954: gap of unknown length  
\* 34955 36580: contig of 1626 bp in length  
\* 36581 36680: gap of unknown length  
\* 36681 37765: contig of 1085 bp in length  
\* 37766 37865: gap of unknown length  
\* 37866 39561: contig of 1696 bp in length  
\* 39562 39661: gap of unknown length  
\* 39662 41262: contig of 1601 bp in length  
\* 41263 41362: gap of unknown length  
\* 41363 43754: contig of 2391 bp in length  
\* 43754 43853: gap of unknown length  
\* 43854 47134: contig of 3281 bp in length  
\* 47135 47234: gap of unknown length  
\* 47235 48919: contig of 1885 bp in length  
\* 48920 49019: gap of unknown length  
\* 49020 51487: contig of 2468 bp in length  
\* 51488 51587: gap of unknown length  
\* 51588 53919: contig of 2332 bp in length  
\* 53920 54019: gap of unknown length  
\* 54020 56113: contig of 2094 bp in length  
\* 56114 56213: gap of unknown length  
\* 56214 58669: contig of 2456 bp in length  
\* 58670 58769: gap of unknown length  
\* 58770 61080: contig of 2311 bp in length  
\* 61081 61180: gap of unknown length  
\* 61181 63791: contig of 2611 bp in length  
\* 63792 63891: gap of unknown length  
\* 63892 66216: contig of 2325 bp in length  
\* 66217 66316: gap of unknown length  
\* 66317 68991: contig of 2675 bp in length  
\* 68992 71765: gap of unknown length  
\* 71766 71865: contig of 2674 bp in length  
\* 71866 73926: gap of unknown length  
\* 73927 74027: contig of 2061 bp in length  
\* 74027 76639: gap of unknown length  
\* 76640 76739: gap of unknown length  
\* 76740 80127: contig of 3388 bp in length  
\* 80128 80227: gap of unknown length  
\* 80228 83868: contig of 3641 bp in length  
\* 83869 83968: gap of unknown length  
\* 83969 88422: contig of 4454 bp in length  
\* 88423 88522: gap of unknown length  
\* 88523 92040: contig of 3518 bp in length  
\* 92041 92140: gap of unknown length  
\* 92141 95492: contig of 3352 bp in length.

FEATURES  
source Location/Qualifiers  
1. .95492  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-11B5"  
BASE COUNT 28940 a 16456 c 17230 g 28330 t 4536 others  
ORIGIN  
Query Match 23.6%; Score 36.6; DB 2; Length 95492;  
Best Local Similarity 58.9%; Pred. No. 2.7; Indels 0; Gaps 0;  
Matches 63; Conservative 0; Mismatches 44;  
QY 8 TCATACCTGTCTTAATAACGGCAGTCATTGACGATGTAGAGTTGCTGGACCTCCAGCAC 67  
||||| || | | | |||| | |||| | |||| | |||| |  
Db 22044 TCATCCCATACCTCCTTCCCAAGGTCAACACAGGATGCTCTATGCTCTCTCA 21985  
QY 68 ACCCCAGGCCCCCAGAGAAGTGGGGCCTCTCTGGTGACCCAGGTTTA 114  
||||| ||||| || | | ||||| || | | ||||| ||  
Db 21984 ACCCCAGGCTCCCTATCACCTGGGGCCTCAAGTCTCTCAAGGGTTA 21938  
Search completed: February 20, 2003, 04:35:08  
Job time : 694.184 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:48:42 ; Search time 3481.23 Seconds  
(without alignments)  
6406.126 Million cell updates/sec

Title: US-09-997-610-1-copy\_2\_1378

Perfect score: 1377

Sequence: 1 atagtgatactactgtctt.....gcctgtacctccattgtatg 1377

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*\*
- 2: em\_esthum:\*\*
- 3: em\_estin:\*\*
- 4: em\_estmu:\*\*
- 5: em\_estov:\*\*
- 6: em\_estpl:\*\*
- 7: em\_estro:\*\*
- 8: em\_htc:\*\*
- 9: gb\_est1:\*\*
- 10: gb\_est2:\*\*
- 11: gb\_htc:\*\*
- 12: gb\_est3:\*\*
- 13: gb\_est4:\*\*
- 14: gb\_est5:\*\*
- 15: em\_estfun:\*\*
- 16: em\_estom:\*\*
- 17: gb\_gss:\*\*
- 18: em\_gss\_hum:\*\*
- 19: em\_gss\_inv:\*\*
- 20: em\_gss\_pln:\*\*
- 21: em\_gss\_vrt:\*\*
- 22: em\_gss\_fun:\*\*
- 23: em\_gss\_mam:\*\*
- 24: em\_gss\_mus:\*\*
- 25: em\_gss\_Other:\*\*
- 26: em\_gss\_pro:\*\*
- 27: em\_gss\_rod:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	675.4	49.0	2615	10	BE420422
2	615	44.7	2009	11	BC004496
3	574.4	41.7	1514	10	BE512633
4	480.6	34.9	1036	13	BM471183
5	479.2	34.8	870	13	BI488505
6	477.8	34.7	898	13	BM457166

7	469.8	34.1	937	14	BQ893011
8	462.6	33.6	1050	13	BM472108
9	451.4	32.8	941	14	BQ723415
10	435.6	31.6	876	14	BQ423563
11	430.4	31.3	717	17	AG086951
12	421.4	30.6	1076	13	BM534723
13	419.8	30.5	676	17	AG064424
14	416.4	30.2	1063	14	BM926985
15	414.8	30.1	1080	14	BQ422247
c	411.4	29.9	654	17	AG112666
16	411.4	29.9	654	17	AG112666
17	411	29.8	721	12	BE728616
18	407.8	29.6	955	12	BE681769
19	405.4	29.4	760	12	BE681182
c	404.6	29.4	736	13	BI754555
20	404.2	29.4	691	17	AG062447
21	404.2	29.4	691	17	AG062447
22	403.8	29.3	900	13	BM449699
23	403.8	29.3	998	14	BM903785
24	402.2	29.2	1182	13	BM451317
25	401.2	29.1	701	17	AQ588873
26	401.2	29.1	704	17	AQ588828
27	399.2	29.0	729	13	BM045873
28	398.4	28.9	720	17	AG116938
29	396.2	28.8	697	17	AG018926
c	394.4	28.6	699	9	AL602279
30	394.2	28.6	833	9	AU140358
31	393.8	28.6	677	17	AG097258
32	393.2	28.6	877	13	BI916691
33	393.2	28.6	659	17	AG092858
34	392.4	28.5	639	17	AG115932
35	390.8	28.4	639	17	AG115932
36	389.4	28.3	676	17	AG125093
37	387.8	28.2	623	17	AQ037711
38	387.8	28.2	704	17	AQ020199
39	386.6	28.1	771	17	AG001229
40	386.4	28.1	885	12	BG761979
41	383.4	27.8	671	17	AG045131
42	381	27.7	742	17	AQ748577
43	380.6	27.6	660	17	AG057341
44	380.2	27.6	686	17	AG119684
45	380	27.6	937	14	BQ223504

ALIGNMENTS

RESULT 1	BE420422	BE420422	2615 bp	mrna	linear	EST 11-OCT-2001
LOCUS	32-393 human bone marrow	cdna library	Homo sapiens	cdna	mrna	
DEFINITION	sequence.					
ACCESSION	BE420422					
VERSION	BE420422.1	GI:16041640				
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	Lu,X.,Cui,L. and Li,Y.					
TITLE	DDRT-PCR from B cell					
JOURNAL	Unpublished (2000)					
COMMENT	Contact: xingwu lu,liangxian cui,yonghai li Department of Biochemistry Institute of Basic Medical Science, Peking Union Medical College DongDan Sanriao 5, Beijing, P.R.C, 100005 Tel: 86-010-85296951 Email: luxingwu@263.net full-length and coding sequence.					
FEATURES	Location/Qualifiers					
Source	1..2615 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="human bone marrow cDNA library" /tissue_type="bone marrow"					

/note="Organ: uterus; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(S). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)." BASE COUNT 705 a 531 c 641 g 738 t

Query Match 49.0%; Score 675.4; DB 10; Length 2615; Best Local Similarity 80.5%; Pred. No. 2.8e-188; Matches 856; Conservative 0; Mismatches 166; Indels 41; Gaps 4;

QY 351 GAATATTGGCTAATGAGGAGCAAAATTTGGCTAATAAGGAAGAAATTTCTAAGCAGCA 410  
 DB 493 GAACCTTGAGCTTGAGAGAGATCATTTAGGGTATCTGGCAGAGAATAATTTCTAAGCAGCA 552  
 QY 411 AAGCATTCAAGAGGTGACTTGGGTGCTGTTAAAGGCAATTCAGTTTCATTAAGGAGGAGCA 470  
 DB 553 AAGCATTTCAAGAGGTAACCTTGGGTGCTGTTAAAGGCAATTTAGTTTATTAAGGGAAGCAGG 612  
 QY 471 GCATTAAGAGTTCAGAAAATTTGCACCCCTGCACAAATGTGATAAAGAAAAGAAAACCCATTTTC 530  
 DB 613 GCATGAAGTTTGGAAAATTTGCAGCCTGCACAAATGCATAGAAAAGAAAATCCCATATTC 672  
 QY 531 TGAGGGGAAATTCAGCTGGCTGCAGAAATTTGCATATGTAATGAGGAGCTGAATGTTAA 590  
 DB 673 TGAGGAGAAATCAAGCTGGCTGCAGAAATTTGCATATGTAACAGGAGTTAAATGTTAA 732  
 QY 591 TCTCAAGCAATGGGAAAATATCTCTGGACATGTCTAGAGGTCTTCACAGAGTCCAT 650  
 DB 733 TCCCAAGACAATGGGAAAATGTCTCCAGGCATGCCAGAGGTCTTCACAGAGCCCT 792  
 QY 651 CAATCACTGGCTCGAGGCCCTAGAG --- AAAATGGTTTGTGGACAGGCCAGGGTC 707  
 DB 793 CCATCACAGGCCCTCGAGGCCCTAGAGGAAAAGTATCTCTGGCAGGCCAGGGTC 852  
 QY 708 CCGTGTCTGTGTGCAGCCCTAGAGACTTGGTGCCCTGTGTCCAGTTAAATTCAGC ----- 761  
 DB 853 CCCATGCTGTGTGCAGCCCTGGGACTTGTGTGCTGTGTCCAGCTGCTCCAGCCATGGC 912  
 QY 762 ----- TGTGGCTTCAGAGGTGTCAGAGGCCCAAGCCCAAGCCTTG 794  
 DB 913 TGAAGAGGGGCCAACATPAGAGCTCGGATTTGGCTTCAGAGGGTGAAGCTGTAAGCCTTG 972  
 QY 795 GCAGCTTCAAGTGTGTGACCTGTGGTCAAGAGAGTCAAGAAATGAGTTTGGCA 854  
 DB 973 GCAGCTTCAAGTGTGTGACCTGTGGTCAAGAGAGTCAAGAAATGAGTTTGGCA 1032  
 QY 855 ACCTCCAATCAGATTTCAGAGATATATGGAACCCCTGGATGCCAGGAGGAAGTTTC 914  
 DB 1033 ACCTCTGCCTAGATTTCAGAGATATATGGAATGCTTGGATGCCCTGGCAGAGTCTGC 1092  
 QY 915 TGTAGGGTGGGTCTCTATGAGAACCTCTCAGAGGTAGTACAAAAGGGAATTTGG 974  
 DB 1093 TGTAGGGAGGGGCCCTCATGAGAAATCTCTGTAGGGCAGTACAGAGGAATTTGGG 1152  
 QY 975 GTGGAGGCCCCACACAGAGTCCCAGTGGGGCTCCATCTAGTACAGCTGTGAGAGAG 1034  
 DB 1153 ATCGAGGCCCCACAGAGTCCCCTACTGGGGACCGCTAGTGGAGCTGTGAGAGAGG 1212  
 QY 1035 TCCAGCATCTCCAGACTCCAGAGGGTATGATCCACTGACAGCTTGCAGCATGTGCTCA 1094  
 DB 1213 GCCACATCTCCAGAGCCCAAGATGATAGATCCACTGACAGCTTGCCTCCATGCACCTGG 1272  
 QY 1095 AAAATCCACAGACTCAGTGCAGCCTGTGAAAGCAGAGGAGTGTGTACCCCTA 1154  
 DB 1273 AAAAACCCGAGA - --- CAACACCCAGCCCATGAAAGCAGCTAGGCGAGGAGCTGTACCCCTG 1328  
 QY 1155 CAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTCTTGCATGTGCATGA 1214  
 DB 1329 CAAGGCCACAGGGGGAGCTGCCCAAGATCATGGGAACCCACCTCTTGCATCAGCATGA 1388

QY 1215 COTGGACGTGAGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTTGACTGCC 1274  
 DB 1389 COTGGATGTGAGATTTGGATTAAGAGGAGATCATTTTGGAGCTTTAAGATTTGACTGCC 1448  
 QY 1275 CACTGATTTGGACATATATGGGCCCGTA - CCCCTTTGTTTGGCCAAATTTTCCAT 1333  
 DB 1449 CTTAGATTTCAGATTCGATGGGCTGTAGCCCTTTGTTTACTCAATTTCTCCCAT 1508  
 QY 1334 TTGGAACCTGCGTATTATTTACCAATGCTGTACCTCCATTTGAT 1376  
 DB 1509 TTGGAACACCTGTGTTACACATACCTGTACCCCATTTGAT 1551

RESULT 2  
 BC004496  
 LOCUS Homo sapiens, Similar to hypothetical protein FLJ14058, clone IMAGE:3831313, mRNA.  
 ACCESSION BC004496  
 VERSION BC004496.1 GI:14709139  
 KEYWORDS HTC.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 2009)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
 REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov  
 COMMENT Contact: MGC help desk  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: ATCC/DCTD/DTF  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Institute for Systems Biology  
 http://www.systemsbio.org  
 contact: amadan@systemsbiology.org  
 Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia Greene, Mark Kettman and Anuradha Madan  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
 Series: IRAL Plate: 14 Row: d Column: 7  
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis  
 This clone has the following problem: frame shifted.  
 FEATURES  
 source Location/Qualifiers  
 1..2009  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3831313"  
 /tissue\_type="Skin, melanotic melanoma."  
 /clone\_lib="NIH\_MGC\_20"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOTB7"  
 BASE COUNT 556 a 429 c 513 g 511 t  
 ORIGIN  
 Query Match 44.7%; Score 615; DB 11; Length 2009;  
 Best Local Similarity 78.2%; Pred. No. 1.9e-170; Matches 781; Conservative 0; Mismatches 205; Indels 13; Gaps 3;  
 QY 391 GAAGAAATTTCTAAGCAGCAAGCATTTCAAGAGTTCAGTTGGTGTCTTAAAGGCATTC 450  
 DB 490 GAAGAAATTTCTATCGCAGCAAGCATTTCAAGAGTTCAGTTGGTGTCTTAAAGGCATTC 549  
 QY 451 AGTTTCATTAAGGAGGAGCAGAGCATTAAGAGTTTCAGAAAATTTGCACCCCTGACATGTGATA 510

Db 550 AGATTATAGGGAGCAGACATAAAAGTTTCAGAAAAATTTGCAGCCTTGACAATGTGATA 609  
Qy 511 AAAAGAAAAACCCATTCTCTGAGGGGAAATTCAGCTGGCTGCAGAAATTTGCATATGT 570  
Db 610 GAGAGAAAAACCCATTCTCTGAGGAGAAATTCAGCCAGCTGCATAAATTTGCATAAGT 669  
Qy 571 AATGAGGAGCTGAATGTTAAATCTCAAGACAATGGGAAATATATCTCTCGACATGTGAC 630  
Db 670 AACAGAGAGCGCATGTTAAATCTCCCAAGACAACGAGAGACATCTCCAGGACATATCAG 729  
Qy 631 AGCTTTCACAGCAGTCATCAATCACTGCTGAGGAGCTGCAGGAGGAGGAGGAGGAGGAGG 690  
Db 730 AGGCTTTCACAGCAGCCTCTCCCTCAGAGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 789  
Qy 691 GGGAC-----AGGCCAGGCTCCTCTGCTGTGTGTCAGCCTAGACAGCTTGGTCC 740  
Db 790 GGACCTGTGCAGCCTAGTACTTGGCACCCTGTGTCCAGCTACTCTCAGTAGTGTCTGAA 849  
Qy 741 CTGTGTCCAGCTAA--TTCAGCTGTGGCTTCAGAGGCTGCAAGCCCAAGCCTTGGCAG 798  
Db 850 AGGGCCCAAGTACAGCTTACGCCATGCTTCAGAGGCTGCAAGCCCAAGCCTTGGCAG 909  
Qy 799 CTTCAAGTGTGTGAGCCTGTGGTGCAGAAAGTCAAGAAATGAGGTTTGGAACT 858  
Db 910 CTTCCGATGCTGTGAGCCTGCAAAATGCACAGAGCTCAAGAAATTTGGATTGGGAACCT 969  
Qy 859 CCAATCAGATTTCAAGAGATATATGGAACCCCTGGATGCCAGCAGAGCTTGTCTCTA 918  
Db 970 CTGCCTAGATTTAGGAGATGTCGGAAGTGTCCGATGCCAGCAAAATTTGCTGCA 1029  
Qy 919 GGGGTGGGCTTCATGAGAACCTCTGCAAGGCTAGTACAAAAGGAAATTTGGGTGG 978  
Db 1030 GGGCAGGGCTTCATGAGAACCTCTGCTAGGCAGTGTGGAAGGGAACATGGGTCT 1089  
Qy 979 GAGCCCCACAGAGTCCCAAGTGGGCTCCATCTAGTAGCTGTGAGAAGAGTCCA 1038  
Db 1090 GAGCCCCACAGAAATCCCTACTGAGGCACACCTAGTGAGCTGTGAGAAAAGAGCCA 1149  
Qy 1039 CCATCTCCAGACTCCAGAGGCTAGTCCACTGACAGCTTGCAGCAGTGCCTGMAAAA 1098  
Db 1150 CCATCTCCACCCAGAGATGATAGATCCACCAACAGCTTGCACCTGCACTGGAANA 1209  
Qy 1099 TCCACAGACTCAGTGCCAGCTGTGAAGCAGCAGGAGTGTGTACCTTACAAA 1158  
Db 1210 GCCGAGACATTAACACACCCGCTGAAGCAGCCAAAGGGGAATATATCTGCAAA 1269  
Qy 1159 ACCGTAGTGAGAGCTGACCAAGCCTGGGAATCTACCTCTTGCATGTGACACCTG 1218  
Db 1270 GCCACAGGGAGAGCTTCCCAAGACTATGGGAACCTACCTCTTACATCAGCAGACCTT 1329  
Qy 1219 GACGTGAGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTTGACTGCCCACT 1278  
Db 1330 TGTGTGAGACCTGGAGTCAAAAGAGATCATTTTGGAGCTTTTAAATCTGACTGGCCCTGCT 1389  
Qy 1279 GGATTTCCGACTTATATGGGGCC--CGTACCCCTTTTGTGGCCAAATTTTTCATTTGG 1337  
Db 1390 GGATTTCCAGACTTCATGGGGCCCTACACCCCTTTGTTTGGCCGATTTCTCCATTTGG 1449  
Qy 1338 AACTGCCGTATTTACCCAAATGCCCTGACCTCCATGTAT 1376  
Db 1450 AAAGCTGTATTTACCCAAATACCTGTACCCCAATGTAT 1488

RESULT 3  
BE512633  
LOCUS BE512633 1514 bp mRNA linear EST 11-OCT-2001  
DEFINITION 32-1514 human B lymphocyte cDNA library Homo sapiens cDNA, mRNA  
sequence.  
ACCESSION BE512633  
VERSION BE512633.1 GI:16041645  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1514)  
AUTHORS Lu,X.,Cui,L. and Li,Y.  
TITLE DRT-PCR from B cell  
JOURNAL Unpublished (2000)  
COMMENT Department of Biochemistry  
Institute of Basic Medical Science, Peking Union Medical College  
DongDan SanJiao 5, Beijing, P.R.C, 100005  
Tel: 86-010-65296951  
Email: luxingwu@263.net  
full cDNA sequence.

FEATURES  
Location/Qualifiers  
1..1514  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="human B lymphocyte cDNA library"  
/tissue\_type="bone marrow"  
/note="Organ: tongue; Vector: pAMP10; mRNA made from  
tongue epithelium, cDNA made by oligo-dT priming.  
Non-directionally cloned into UDG sites. Size-selected on  
agarose gel, average insert size 500 bp. Primary library.  
cDNA Library Preparation: David B. Krizman, Ph.D.  
REFERENCE: Krizman et al. (1996) Cancer Research  
56:5380-5383."

BASE COUNT 380 a 341 c 398 g 395 t  
ORIGIN

Query Match 41.7%; Score 574.4; DB 10; Length 1514;  
Best Local Similarity 81.0%; Pred. No. 1.8e-158;  
Matches 734; Conservative 0; Mismatches 131; Indels 41; Gaps 4;

Qy 508 ATAAAAAAGAAAAACCCATTCTCTGAGGGGAAATTCAGCTGGCTGCAGAAATTTGCATA 567  
Db 1 ATAGAAAAAATATCCCATATCTCGAGGAGAAATTCAGCTGGCTGCAGAAATTTGCATA 60  
Qy 568 TGTATGAGGAGCTGAATGTTAAATCTCAAGACAATGGGAAATATCTCTCGGACATGT 627  
Db 61 TGTACAGAGGATTAATGTTAAATCTCCCAAGACAATGGGAAATGCTCTCCAGGCGCATGC 120  
Qy 628 CAGAGGCTTTCACAGCAGCTCCATCAATCACTGGCTTGAGGCTTAGAG---AAAAAGG 684  
Db 121 CAGAGGCTTTCACAGCAGCCCTCCCATCAGAGGCTTCAGGCTTAGGAGGAAAAAGTGA 180  
Qy 685 TTTTGTGGCAGAGCCAGGCTCCTGTGTGTGTGAGCCTTAGAGACTTGTGCGCCTGT 744  
Db 181 TCTCTTGGGCCAGGCCAGGCTCCCATGCTGTGTGAGCCTTGCGGACTTGTGCTTGT 240  
Qy 745 GTCCAGTAAATTCAGC-----TGTCGCTTCA 771  
Db 241 GTCCAGCTGTCTCCAGCCATGGCTGAAAGGGCCCAACATAGAGCTCGGATGTGCTTCA 300  
Qy 772 GAGGTGCAAGCCCCCAAGCCTTGGCAGCTTCCAAATGCTGTGTGAGCCTTGTGGTGCAAG 831  
Db 301 GAGGTGCAAGCTGTAGCCCTTGGCAGCTTCCATGCTGTGTGAGCCTGCCAGTGCACAG 360  
Qy 832 AAGTCAAGAAATGAGGTTTGGGAACCTCCAATCAGATTTTCAGAAATATATAGGAACCCC 891  
Db 361 AAGTCAAAATTTGGGGTTTGGGAACCTCTGCTAGATTTTCAGAAATGATATGGAATGCC 420  
Qy 892 TGGATGCCCAGCAGAAAGTTTGT 951  
Db 421 TGGATGCCCTGGCAGAGCTGT 480  
Qy 952 GTAGTACAAAAGGGAATTTGGGTGGGAGGCCCCACACAGAGTCCCGAGTGGGCTTCCA 1011  
Db 481 GCAGTACAGAGGGAATTTGGGATCGGAGCCCGCCAGCAGACAGATCCCTACTGGGCGACCG 540  
Qy 1012 TCTAGTAGAGCTGTGAGAAGAGTCCACCATCTCTCCAGACTCCAGAGGGTAGATCCACT 1071  
Db 541 CCTAGTGGAGCTGTGAGAAGAGGGGCCACCATCTCTCCAGAGGAGGAGGAGGAGGAGGAGG 600



AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM11479 row: k column: 04  
High quality sequence stop: 849.  
Location/Qualifiers  
1. .870  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5191803"  
/clone\_lib="NIH\_MGC\_114"  
/lab\_host="DH10B"  
/note="Organ: Brain; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 019. Note: this is a NIH\_MGC Library."  
BASE COUNT 222 a 211 c 256 g 181 t  
ORIGIN  
Query Match 34.8%; Score 479.2; DB 13; Length 870;  
Best Local Similarity 76.1%; Pred. No. 2.3e-130;  
Matches 645; Conservative 0; Mismatches 168; Indels 35; Gaps 3;  
QY 492 GCACCCTGACAAATGTGATAAAGAAAAACCCATTTTCTGAGGGAAATTCAGAGCTGGC 551  
DB 1 GCAGCCTGATCATGAAGTAGAAGAAAAACCCATTTTTCAGGATAAATTCAGAGCCAGC 60  
QY 552 TGCAGAAATTTGCATATGTAATGAGGAGCTGAATGTTAATCCTCAAGACAATGGGGAAAA 611  
DB 61 TGCAGAAATTTGCATATGTAAGCAGAGTCTAATGTTAATCTCCAAGACCACTGGGAAAA 120  
QY 612 TATCTCTGGACATGTCAGAGGCTCTCACAGCAGCTCCATCAATCACTGGCCTGGAGGC - 670  
DB 121 TGTCTCCGGCCATGTCAAGACCTTCATGGCAGCCCTCCCATCAAGAGGCTGGAGGCCA 180  
QY 671 CTAGGAGAAATGTTTTGTGGGACAGGCCAGGCTCCCTGTGCTGTGTGAGCCTAGAG 730  
DB 181 GGAGAAAAAGAGGTTTCGTGGGCGAGGCCAGCATCCCGTGTCTGTGAGCCTAGG 240  
QY 731 ACTTGTGCTGTCCTCCAGTTAATCA----- 759  
DB 241 ACTTGTGCTGTCCTCCAGCTCTCCAGCTGTGACTGAAAGGGGCCAAGTAGAGTCC 300  
QY 760 --GCTGTGGCTTCAGAGGTCAGAGCCCAAGCCCTGCGAGCTTCCAGTGTGCTGAGC 817  
DB 301 GGGCTGTGGCTTCAGAGGTTGGAAGCTCTAAGCTTTTGGCAGCTTCCAGTGTGCTGAGC 360  
QY 818 CTGTGGGTGCAAGAAGTCAAGAAATTTAGGTTTGGGAACCTCCCAATCAGATTTTCAGAAGA 877  
DB 361 CTACAGGTGCACAGACTCAAGAACTGAGGTTTGGGAACCTCTCCCTAGATTTTCAGAAGA 420  
QY 878 TATATGGAACCCCTGGATGCCAGGAGGAGTTTGTGTGTAGGGTGGGGTCTCATGGA 937  
DB 421 TGTTTGGAATTCGCTGGATGCCAGGAGGAGTTTGTGTGTAGGGTGGGGTCTCATGGA 480  
QY 938 GAACCTCTGCAAGGTTAGTCAAAAGGGAATGTTGGTGGGAGCCGCCACACAGAGTCC 997  
DB 481 GAACCTCTGCTGGAGCAGTCAGAGAAAGAAATGTGGGATCGGAGCCGCCACAGAGAGTCC 540

QY 998 CCAGTGGGGCTCCATCTAGTAGAGCTGTGAGAAGAGTCCACCATCTCCAGACTCCAGA 1057  
DB 541 ATACTGGGGCCTCCTAGTAGTGAACCTGTGAGAAGACACACCCGCTCTCCAGACCCCA 600  
QY 1058 AGGTAGATCCACTGACAGCTTGCGAGCATGTGCCTGAAAAATCCACACACACTCAGTGCC 1117  
DB 601 ATGGTAGATCCCAACACAGCTTGCCATGTGCTGGAAAAAGATGCCAGTACTCAATGCC 660  
QY 1118 AGCCTGTGAAAGCAGCAGGAGTGGAGTCTGTACCTTACCAAAACCGTAGTGGCAGAGCTGA 1177  
DB 661 AGCCCATGAAGGCAGCCAGGAGGAGGCTGTACCTCTGCAAGCCACAGGGAGGAGCTGT 720  
QY 1178 CCAAGACCGTGGGAATCTACCTCTTGCATTTGCATGACCTGGAGCTGAGAGATGAGTCA 1237  
DB 721 CCAAGACCATGGAAACCCACTCTTGCATCAGCGTGACCTGGATGCGAGAGCTGGAGTCA 780  
QY 1238 AAAGAGATCAATTTGGAGCTTTAA-GATTTGACTGCCCACTGGATTTTCGAGCTTATATG 1296  
DB 781 AAGGAGATCAATTTGGAGCTTTAAACATGTGACCTGCGCTGCTGGATTCAGACTTGCAT 840  
QY 1297 GGGCCCGT 1304  
DB 841 GGGGCCCT 848  
RESULT 6  
BM457166  
LOCUS BM457166  
DEFINITION AGENCOURT\_6411690 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:5583427  
5', mRNA sequence.  
ACCESSION BM457166  
VERSION BM457166.1 GI:18506206  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 898)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM12346 row: h column: 20  
High quality sequence start: 9  
High quality sequence stop: 713.  
Location/Qualifiers  
1. .898  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5583427"  
/clone\_lib="NIH\_MGC\_92"  
/tissue\_type="embryonal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: Testis; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."  
BASE COUNT 225 a 228 c 254 g 191 t  
ORIGIN

Query Match 34.7%; Score 477.8; DB 13; Length 898;  
Best Local Similarity 77.3%; Pred. No. 6.1e-130;  
Matches 669; Conservative 0; Mismatches 137; Indels 59; Gaps 5;

```
QY 562 TCATATGTAATGAGAGCTGAATGTTAATCCTCAAGACAATGCGGAAAAATATCTCCTGG 621
Db 35 TGCATAAGCAGCAGGAGCTAATGTTAATCCTCAAGACAATGCGGAAAAATGCTCCAG 94
QY 622 ACATCTCAGAGGCTCTCAGACAGTCCATCAATCACTGGCTGGAGCCCTAGGAG---A 678
Db 95 TCATGTCAGAGACCTTCATGGCAGAGCCCTCCCATCACAGGCCAGAGCCCAAGAGAAA 154
QY 679 AAATGGTTTTTGGGACAGCCCGAGGCTCCCTGCTGTGTGTCAGCCCTAGAGACTGGTG 738
Db 155 AAGTGGTTTTTGGGCTGGGCCAGGCTCTGCTGTGTGTCAGCCCTAGGAGACTGGTG 214
QY 739 CCTGTGTCCTCAGTTAATCA-----CCTGTG 765
Db 215 CCTGTGTCCTCAGCTCTCCAGCCATGGCTGAAGAGGCAACATAGAGCTTGGGCTGTG 274
QY 766 GCTTCAGAGGCTGCAAGCCCTGGCAGCTTCCAAAGTGGTGTGAGCCTGTGGT 825
Db 275 GCTTCAGAGGCTGCAAGCTTGGCAGCTTCCAAAGTGGTGTGAGCCTGTGGT 334
QY 826 GCAAGAGAGTCAAGAAATGAGGTTTGGGAACCTCCCAATCAGATTTTCAGAGATATATGA 885
Db 335 GCAGAGAGTCAAGAAATGAGATTTGGGAACCTCTGCTAGATTTTCAGAGATATATGA 394
QY 886 AACCCCTGGATGCCAGGAGGTTTCTGTGTGAGGCTGCTCATGGAGAACCTCT 945
Db 395 AACACCTGGATGCCAGGAGGTTTCTGTGTGAGGCTGCTCATGGAGAACCTCT 454
QY 946 GCAAGGCTAGTCAAAAGGAAATGTTGGTGGGAGCCCAACAGAGTCCCAAGTGGG 1005
Db 455 CCTAGGAGGAGTGTGAAGGGAATG-TGGGGCAGAGAGCCCAACAGAGTCCCAAGTGGG 513
QY 1006 GCTCCTAGTGTAGAGCTGTGAGAGAAAGTCCACCATCCTCC----- 1047
Db 514 GCAGCTGCCTAGTGTGAGCTGTGAGAGAGAGCCCAAGTCCCAAGAGCTGAGCCCAAGAT 573
QY 1048 ---AGACTCCAGAGGGTAGATCCACTGACAGCTTGCAGCATGTGCCTGCAAAATCCACA 1104
Db 574 GGTAGACCCCAAGATGGTAGATCAACCAACCACTTGTGCTTGTGCTGGAAGACTACA 633
QY 1105 GACACTAGTGCCAGCCTGTGAAAGCAGCAGGGATGTGATGCTGTACCTACAAACCGTA 1164
Db 634 GACACTCAACATCAGCCTGTGAAAGCAGCAGGGAGGCTGTACTCTGCAAAAGCCACA 693
QY 1165 GTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTCTTGCATTTGCATGACCTGGAGCTG 1224
Db 694 GGGGGGAGCTGCACAGACCATGGACCTATCTCTTGCATCAGCATGACCTGATGTG 753
QY 1225 AGACATGGAGTCAAAAGAGATCAATTTGGAGCTTTAAGATTTGACTGCCCACTGGATTT 1284
Db 754 AGACATGGAGTCAAAAGAGATCAATTTGGAGCTTTAAGATTTGACTGCCCACTGGATTT 813
QY 1285 CGGACTTATATGGGGCCGCT-ACCCTTTGTTTGGCCCAATTTTCCATTTGGAACCTGC 1343
Db 814 CAGACTTCGATGGACCCCTGAACCCCTTTGTTTGGCCCAATTTCTCTCTTTGGAACAGC 873
QY 1344 CCTATTATACCAATGCTGTACCTC 1368
Db 874 TGTATTACCAATACCTGTACCC 898

RESULT 7
BO893011
LOCUS BO893011 937 bp mRNA linear EST 16-AUG-2002
DEFINITION AGENCOURT_8122304 Lupski_dorsal_root_ganglion Homo sapiens cDNA
clone IMAGE:6179271 5', mRNA sequence.
ACCESSION BO893011
VERSION BO893011.1 GI:22285025
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE 1 (bases 1 to 937)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13560 row: c column: 16
High quality sequence stop: 722.
FEATURES
source
Location/Qualifiers
1..937
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6179271"
/clone_lib="Lupski_dorsal_root_ganglion"
/sex="male"
/tissue_type="dorsal root ganglia"
/dev_stage="adult, 36 yr"
/lab_host="DH10B"
Note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGCGTCCG-3' and
5'-TCGACTGTTCTAGATCGCAGCGCCGCT(15)-3'. Size selected >
1 kb for average insert length 1.7 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
BASE COUNT 235 a 226 c 258 g 214 t 4 others
ORIGIN
Query Match 34.1%; Score 469.8; DB 14; Length 937;
Best Local Similarity 75.4%; Pred. No. 1.4e-127;
Matches 682; Conservative 0; Mismatches 149; Indels 74; Gaps 5;
QY 514 AAGAAAAACCCATTTCTTGAGGGGAAATTCAGCTGGCTGCAGAAATTTGCATATGTAAT 573
Db 3 ATGAAATTTGCATTTCTTGAATAGAAATTCAGCCAGCTGCAGAAATTTGCATAAGGAAG 62
QY 574 GAGGAGCTGAATGTTAATCCTCAAGCAATGGGAAATATCTCTGGCATGTGAGAGG 633
Db 63 GAGGAGCAAAATGTTAATCCCAAGACAATGGGAAATGTCTCCAGGCCATGTGAGAGG 122
QY 634 TCTTTCACAGCAGTCCATCAAAATCACTGGCTGGAGCCCTAGGAG---AAAATGGTTTGT 690
Db 123 CTTTCAGGCGAGTCCCTCCCATCAGGCCCAAGAGCCCTAGGAGGAAATGTTGTTTCT 182
QY 691 GGGACAGGCCCGAGGTCCTGTGCTGTGTGACGCTTAGAGACTTGTGCTGCTGTGTCCTCA 750
Db 183 GGGCTGGGCCCGAGGTCCTCAAGCTATGTGACGCTT-GGGGTTTGTGCTCCACTGCATCCA 241
QY 751 GTTAATTCAGCTG-----TGCTTCAGAGGCT 777
Db 242 GCTGCCCGAGCTGTGGCTGAAGGGGCAATAGAGCTCGGGCCCTGCTGTTTAGATGCT 301
QY 778 GCAAGCCCAAGCCCTTG-----GCAGCTTCCCAAGTGGTGTGAGCCTGGGTGCAAG 831
Db 302 GCAAGCCCAAGCCCTGTGGTGTGAGCTTCCATGTGGTGTGAGCCTGAGCTGCAGCTGCACAG 361
QY 832 AAGTCAAGAATTTAGGTTTGGGAACCTCCAATCAGATTTTCAGAGATATATGGAACCC 891
Db 362 AAGTCAAGAATTTAGGTTTGGGAACCTCCACCTAGATTTTCAGAGATATGTAAGAAATGCC 421
QY 892 TGGATGCCCGAGCAGAAATTTGCTGTAGGGTGGGCTCCTCATGAGAACCTCTGCAAGG 951
Db 422 TGGATGCCCGAGCAGAAATTTGCTGCAAGGGTGGGCTCCTCATGAGAACCTCTGCTAGG 481
```



ACCESSION BQ723415  
 VERSION BQ723415.1 GI:21862312  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 941)  
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Dr. James R. Lupski  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: LLAM13596 row: b column: 17  
 High quality sequence stop: 606.  
 Location/Qualifiers  
 1..941  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6193072"  
 /clone\_lib="Lupski\_sympathetic\_trunk"  
 /sex="male"  
 /tissue\_type="sympathetic trunk"  
 /dev\_stage="adult, 16 yr"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6 (Life Technologies); Site\_1: NotI; Site\_2: SalI; cDNA made by oligo-dT priming. Directionally cloned using the following adaptors: 5'-TCGACCCAGGCTCG-3' and 5'-GACTAGTTCTAGATGCGAGCGCGCCCT(15)-3'. Size selected > 1 kb for average insert length 1.9 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine); available through Life Technologies."  
 BASE COUNT 257 a 208 c 267 g 209 t  
 ORIGIN  
 Query Match 32.8%; Score 451.4; DB 14; Length 941;  
 Best Local Similarity 78.0%; Pred.No. 4.1e-122;  
 Matches 672; Conservative 0; Mismatches 146; Indels 43; Gaps 9;  
 QY 391 GAAGAAATTTCTAAGCAGCAAGCATTCGAAGGTGACTTGGGTGCTGTAAAGGCATTC 450  
 DB 51 GAAGAAATTTCTAAGCAGCAAGCATTCGAAGGTCTCTTGGGTGCTGTAAAGCATTC 110  
 QY 451 AGTTTCATAGGGAGCAGACATAGAGTTCAGAAAATTTGCACCCCTGACAAATGTGATA 510  
 DB 111 AGTTTAAAGGAAACAGCAGCAATAAAGTTTGGAAAATTTGCAGCTGCAACAATGCA 170  
 QY 511 AAAAGAAAACCCATTTCTCAGG- GGAAATTCAGCTGGCTGCAGAAATTTGCATATG 569  
 DB 171 GAAAGAAAATCCCATTTTCTGAGGAGAAAGTCAAGCTGGCTGCAGAAATTTACATAG 230  
 QY 570 TAATGAGGAGTGAATGTTAATCCTCAAGCAAT- GGGGAAAATATCTCCTGGACATGTC 628  
 DB 231 TAACAAGAGCAATGTTAATCCCAAGCAATGGGGTAATGTCTCCAGAGCATATC 290  
 QY 629 AGAGGTCTTCACAGCAGTCCATCAATCACTGGCCCTGGAGGCGCTAGGAG- ---AAAATGTT 685  
 DB 291 AGAGGTCTTCACGGGAGCGCCCTCCCTTTACAGGCTGGGGGCTAGGAGGAAAAGTGGT 350  
 QY 686 TTTTGGGACAGCCAGGCTCCCTGCTGCTGTCAGCCTAGACAGTGGTGGCCCTGTG 745  
 DB 351 TTCGTGGGCTGACACAGAGGGTCCCATGCTGTGTGCAGCCTAGGAGACTTGGTGGCCCTGCA 410

QY 746 TCCAGTTAAATTC-----GCTGTGGCTTCAG 772  
 DB 411 TCCAGTCTCCAGCTGTGGCTGAAAGGGGCTATCACAGAGCTCAGGCTGTGGCTTCAG 470  
 QY 773 AGGGTGCAGCCCAAGCCTTGGCAGCTTCCAAGTGGTGTCTGAGCCTGTGGGTGCAAGA 832  
 DB 471 AGGGTGCAGTCCCAAGCCTTGGCAGCTTCCAAGTGGTATTGAGCCTGCAAGTACATGA 530  
 QY 833 AGTCAAGAAATTGAGTGTGGGAACCTTCCAATCAGATTTTCAGAAGATATATGAAACCCCT 892  
 DB 531 AGTCAAGAAATTGAGTGTGGGAACCTTCCAATCAGATTTTCAGAAGTACGGAATACCT 590  
 QY 893 GGATCCCGCAGCAGAAAGTTTGTAGGGTGGGTGCTTCATGAGAGACCTCTGCAAGGG 952  
 DB 591 GGATCCCGCAGCAGAAAGTTTGTAGGGTGGGTGCTTCATGAGAGACCTCTGCTAGAG 650  
 QY 953 TAGTACAAAAGGAAATTTGGTGGGAGCCCGCCACACAGAGTCCCGAGTGGGGCTC-CA 1011  
 DB 651 TAGTGGGGAGGAAATTTAGGGTTCAGAGCCCGCCACACAGAGTCCCTACTGAGACACTGC 710  
 QY 1012 TCTAGTAGAGCTGTGAGAAAGTCCACCATCC-TCCAGACTCCAGAAGGGTAGATCCAC 1070  
 DB 711 CCTAGTGGAGCTGTGAGAAAGCCATCATCTTTTCAGACCCAGGAATGGTAGATCCAC 770  
 QY 1071 TGACAG-CTTGCAGCATGTGCTGAAATTCACAGACACTCAGTCCCGCAGCTGTGAAAG 1129  
 DB 771 TGACAGCTTGCACCTGTGTGCTGGAAAGCCACAGACACTCAACGCCAGCCCATTAAG 830  
 QY 1130 CA-GCAGGATGGAGTGTACCCCTACAAAACCGTAGTGGGAGAG-CTGACCAAGACCT 1187  
 DB 831 CACCAGGGAGGGAGGCTGTACCCCTGCAAAAGCCACAGAGGGTGGGACAGCCCAAGACTAT 890  
 QY 1188 GGAATCTACCTTTGCAATTG 1208  
 DB 891 GGGACCCACCTCTTTTGTG 911  
 RESULT 10  
 BQ423563  
 LOCUS AGENCOURT\_7845919 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:6064274  
 DEFINITION 5', mRNA sequence.  
 ACCESSION BQ423563  
 VERSION BQ423563.1 GI:21118878  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 876)  
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: ATCC/DCTD/DTF  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: LLAM13338 row: 1 column: 03  
 High quality sequence stop: 674.  
 Location/Qualifiers  
 1..876  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6064274"  
 /clone\_lib="NIH\_MGC\_72"  
 /tissue\_type="melanotic melanoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="organ: skin; Vector: pCMV-SPORT6; Site\_1: NotI;

FEATURES  
source



```
||||| 570 AGACATGGAGCCAAAGGAGATCATTTGGAGCTTTAAAGATTTGACTTCCCTGCTGGATTT 629
QY 1285 CGGACTTATATGGGGCCGCTA--CCCTTTGTTTGGGCAATTTTTCATTTGGAACGTG 1342
Db 630 CAGACTTGCAATGGGGCCCTGTAGCCCTTTTGTGGGCAATTTCCCAATTTGGAATGG 689
QY 1343 CGGTATTACCCAA 1356
Db 690 GGTAAATTTACCCA 703

RESULT 12
BM554723 1076 bp mRNA linear EST 20-FEB-2002
AGENCOURT_6578521 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5468842
5' mRNA sequence.
ACCESSION BM554723
VERSION BM554723.1 GI:18794564
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DP
cDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1973 row: b column: 11
High quality sequence stop: 665.
FEATURES
Source
1..1076
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cclone="IMAGE:5468842"
/clone_lib="NIH_MGC_41"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH-MGC Library."
BASE COUNT 274 a 251 c 295 g 256 t
ORIGIN
Query Match 30.6%; Score 421.4; DB 13; Length 1076;
Best Local Similarity 78.2%; Pred. No. 3.2e-113;
Matches 559; Conservative 0; Mismatches 121; Indels 35; Gaps 3;
QY 697 GCGCCAGGTCCTGTGTGTGTGTCAGAGCTTAGAGACTTGGTCCCTGTGTCCAGTTAAT 756
|||||
Db 5 GCGCCAGGTCCTGTGTGTGTACAGCTAGGACTGGTGTCTGTGTCCAGCCACT 64
QY 757 TCAG-----CTGTGGCTTCAGAGGTCAGGC 783
|||||
Db 65 CCAGCAGTGGTGAAGGACCAACATAGAGCTCGACTGTGGCTTCAGAGAGTCCAAGC 124
QY 784 CCCAGCCTTGACGCTCCAGTGGTGTGAGCCTGTGGGTGCACAAAGTCAAGATT 843
Db 125 CCCAGTCTGGCAGCTTCCACATGGTGTGAGCCTGCAAGTGCACATAAAGTCAAGATT 184
```

```
QY 844 GAGGTTTGGGAACCTCCATCAGATTTTCAGAAATATATGGAACCCCTGGATGCCAGG 903
|||||
Db 185 GGAGTTTGGAACCTCCGCTAGATTTTCAGAGGATGTATGGAATGCCCTGGATGCCAGG 244
QY 904 CAGAAGTTTGTGTAGGGTGGGTCCTCATGTGAGAACCTCTGCAAGGTTAGTACAAAAG 963
|||||
Db 245 AAGAAGTGTGCTACAGGGCAGGCGACTCATGGAGAACCTCTGCTAGGACATGCCAGAAG 304
QY 964 GGAATGTTGGTGGAGCCCCACACAGAGTCCCCAGTGGGGCTCCATCTAGTAGAGCT 1023
|||||
Db 305 GGAATGTTGGGTGAGAGCCCCACACAGAGTCCCTCATTTGGGGCACTGCCTAGTGAGGT 364
QY 1024 GTGAGAAGATGTCACCATCTCCAGACTCCAGAGGGTAGATCCACTCAGAGCTTGCAG 1083
|||||
Db 365 ATGAGAAGAGGCCATCATCTACTCGAGACCCAGCATGTTAGATCCACCCAGACTTGCAC 424
QY 1084 CATGTGCTGAAAAATCCACAGACTCAGTCCGACCTGTGAAAGCAGCAGGGATGGAG 1143
|||||
Db 425 TGTGCACCTGGAAAGCCACAGACTCAATGCCAGCCCATGAAAGCCGCTGAGAGGGAG 484
QY 1144 TCTGTACCTTACAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGATCTACCTCTTG 1203
|||||
Db 485 GCTGTACCTTACAAAAGCCACAGGAGAGTCCGCCAGGCCATGTGAACCCACTTTCG 544
QY 1204 CATTTGCTGACCTGGAGCTGAGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGA 1263
|||||
Db 545 CATCAGCATGACTCGGATGTGAGACATGGAGTCAAAAGGAGATAATTTGGAGCTTTAAGA 604
QY 1264 TTTGACTGCCCCACTGGATTTTCGGACTTATATGGGGCCCGTA-CCCTTTTGTGTTGGCCA 1322
Db 605 TTTGACTGCCCCACTGGATTTTCGGACTTGTCTGCTGCTAGCCCTTTGTTTGGCCA 664
QY 1323 ATTTTTCAT--TTGGAAGTCCGCTATTACCAATGCTGTACCTGCCATTCAT 1376
|||||
Db 665 ATTTCTCATGTGGGAATGGCTGTATTACTCAATGCTGAATCCCCATTTGAT 719

RESULT 13
AG064424 AG064424 676 bp DNA linear GSS 03-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-053H24.R, genomic survey sequence.
ACCESSION AG064424
VERSION AG064424.1 GI:16616226
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC Library clone:PTB-053H24.R.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 676)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbpes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector : pKSI45
R.Site 1 : SacI
R.Site 2 : SacI.
```



```
QY 1304 -TACCCCTTTGTTGGCAATTTTTCATTGGAAGTCCGCTATTACCAATGCTTG 1362
|||||
Db 609 CAACCCCTTTGTTGGCAATTTCTCCATTGGAATGGCTGTATTACCCCAATACCTG 668

QY 1363 TACCTCCATTGTAT 1376
|||||
Db 669 TACCCCCCATGTAT 682

RESULT 15
BQ422247/c
LOCUS BQ422247 1080 bp mRNA linear EST 23-MAY-2002
DEFINITION AGENCOURT_7766790 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6015590
5', mRNA sequence.
ACCESSION BQ422247
VERSION BQ422247.1 GI:21117562
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1080)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM13211 row: o column: 15
High quality sequence stop: 702.
FEATURES
source
1. 1080
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6015590"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dr primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 239 a 304 c 221 g 316 t
ORIGIN
```

Query Match 30.1%; Score 414.8; DB 14; Length 1080;  
Best Local Similarity 79.9%; Pred. No. 2.9e-111;  
Matches 566; Conservative 0; Mismatches 102; Indels 40; Gaps 5;

```
QY 390 GGAAGAAATTTCTAGACGACCAAGCATTTCAAGAGTGACTTGGGTGCTGTTAAAGGCATT 449
|||||
Db 708 GGAAGAAATTTCTAGACGACCAAGCATTTCAAGAGTGACTTGGGTGCTGTTAAAGGCATT 649

QY 450 CAGTTTTCATAAGGGAGGACGACATAAGAGTTTCAGAAAATTTGCACCTCGACAATGTGAT 509
|||||
Db 648 CAGTTTTCATAAGGGAGGACGACATAAGAGTTTCAGAAAATTTGCACCTCGACAATGTGAT 589

QY 510 AAAAAAGAAAAA-CCATTTTCTGAGGGAAATTCAGAGCTGGCTGCAGAAAATTTTCATAT 568
|||||
Db 588 AGAAAAGAAAAATCTCATTTTCTGAGGAGAAATGGAAGCAGGCTGCAGAAAATTCGCCATA 529

QY 569 GTAATGAGGAGCTGATCTTAATCTCTAAGACAAATGGGGAAAAATATCTCTGGACATGTC 628
|||||
Db 528 GTAATGAGGAGCCTAATCTTAATCTCTAAGACAAATGGGGAAAAATATCTCTGGGATATGTC 469
```

```
QY 629 AGAGGTCTTTCACAGCAGTCCATCAATCACTGGCCTGGAGGCT---AGAGAAAAATGGT 685
|||||
Db 468 AGA-GTTTTTCATGGCAGCCCTTCCATCACAGGCCACAGAGGCTTACAAGAAAAATGGT 410

QY 686 TTTGTGGGACAGGCCACAGGGT--CCCTGTGCTGTGTGTCAGCCTTAGAGACTTTGGTGCCCTG 743
|||||
Db 409 TTTCTGGGCCAGGCCACAGGGTACACTCGTGTGCTGTGTCAGCAAGAGACTTTGGTGCTCTG 350

QY 744 TGTCCCAAGTTAATTCA-----GCTGTGCTTTC 770
|||||
Db 349 CATCCCAAGTCGCTCCAGACATGGCTGAAAAGGGGCCAACATAGAGTTTCGGGCAGTGGCTTC 290

QY 771 AGAGGTGCAAGCCCAAGCCCTTGGCAGCTTCCAAAGTGTGTGAGCCTGTGGGTGCAAA 830
|||||
Db 289 ACAGGGTGCAGGCCCAAGCCCTTGGCAGCTTCCATGTGTGTGAGCCTTGAGAGTGCACA 230

QY 831 GAAGTCAAGAATTGAGGTTTGGGAACTTCCAAATCAGATTTTCAGAAAGATATATGGAACCC 890
|||||
Db 229 AAAGTCAAGAAGTGGGGTTTGGGAACCTCTGCTAGATTCTGAAGATGTATGGAATGC 170

QY 891 CTGGATGCCAGGCAGCAAGCTTTTCTGTAGGGTGGGGTCTCTCATGGAGAACCTCTGCAAG 950
|||||
Db 169 CTGGATGCTCAGGCATTAAGTTTGTGTCAGGGCAGGGCTCTCATGGCGAACCACTTTAG 110

QY 951 GGTAGTACAAAAGGGGAAATGTTGGGTGGAGGCCCCACACAGAGTCCCCAGTGGGGCTCC 1010
|||||
Db 109 GGCAGTGGAGAGGAAATGTGAGGTTCAGAGCCCCCACACAGAGTCCCCCTGGGGCACC 50

QY 1011 ATCTAGTAGAGCTGTGAGAAAGTCCACCATCTCTCCAGACTCCAGAA 1058
|||||
Db 49 ACCTAGTGGAGCTGTGAGAAAGAGGGCCACTGTCTCTCCAGACCTCAGAA 2
```

Search completed: February 20, 2003, 06:15:08  
Job time : 3510.23 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 18:01:21 ; Search time 110.805 Seconds  
(without alignments)  
6329.459 Million cell updates/sec

Title: US-09-997-610-1\_COPY\_2\_1378

Perfect score: 1377

Sequence: 1 atagtgtcatacctgtctt.....gcctgtacctccattgtatg 1377

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 424239 seqs, 254661826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1377	100.0	1381	9	US-09-997-610-1
2	1377	100.0	1731	9	US-09-997-610-5
3	988.4	71.8	1377	9	US-09-997-610-3
4	988.4	71.8	1731	9	US-09-997-610-7
5	641.6	46.6	155074	9	US-10-026-188-6
6	618.2	44.9	465237	10	US-09-933-267A-1
7	610.2	44.3	180557	12	US-10-003-806-6
8	610.2	44.3	180557	12	US-10-003-806-9
9	603.4	43.8	170834	10	US-09-835-232-7
10	590.4	42.9	202001	10	US-09-734-674-3
11	553.8	40.2	684973	10	US-09-263-959-1
12	545.2	39.6	69327	10	US-09-777-921A-3
13	523.6	38.0	25603	9	US-09-819-607-3
14	523	38.0	1946	10	US-09-864-761-2925
15	502.8	36.5	148567	9	US-10-254-869-3
16	502.8	36.5	148567	9	US-09-801-876B-3
17	469.6	34.1	180557	12	US-10-003-806-6
18	469.6	34.1	180557	12	US-10-003-806-9
19	463	33.6	197997	10	US-09-822-246-3

c	20	450.8	32.7	143068	10	US-09-967-768A-316	Sequence 316, App
c	21	439.6	31.9	49136	10	US-09-768-877-1	Sequence 1, Appli
c	22	434.8	31.6	11469	10	US-09-764-877-2791	Sequence 2791, Ap
c	23	430.8	31.3	761	10	US-09-864-761-19706	Sequence 19706, A
c	24	420.2	30.5	15819	10	US-09-764-877-2765	Sequence 2765, Ap
c	25	417	30.3	12104	10	US-09-764-877-2761	Sequence 2761, Ap
c	26	400.2	29.1	1666	9	US-10-001-835-76	Sequence 76, Appl
c	27	398.2	28.9	55795	10	US-09-880-107-1543	Sequence 1543, Ap
c	28	392	28.4	577	10	US-09-864-761-6870	Sequence 6870, Ap
c	29	391.6	28.4	3816	10	US-09-880-107-2298	Sequence 2298, Ap
c	30	391.2	28.4	7642	10	US-09-764-877-2764	Sequence 2764, Ap
c	31	382	27.7	11071	10	US-09-764-877-2793	Sequence 2793, Ap
c	32	370.2	26.9	42999	10	US-09-740-029-3	Sequence 3, Appli
c	33	369.4	26.8	63588	9	US-10-243-735-3	Sequence 19812, A
c	34	362	26.3	3693	10	US-09-864-761-19812	Sequence 3941, Ap
c	35	356.6	25.9	2093	10	US-09-880-107-3941	Sequence 7, Appli
c	36	356.4	25.9	170834	10	US-09-835-232-7	Sequence 5964, Ap
c	37	347.4	25.2	594	10	US-09-864-761-6964	Sequence 6964, Ap
c	38	341.2	24.8	573	10	US-09-864-761-9752	Sequence 9752, Ap
c	39	327	23.7	847	10	US-09-764-887-623	Sequence 623, App
c	40	311.6	22.6	575	10	US-09-864-761-8786	Sequence 8786, Ap
c	41	311.4	22.6	489	10	US-09-864-761-5584	Sequence 5584, Ap
c	42	304.8	22.1	578	10	US-09-864-761-9187	Sequence 9187, Ap
c	43	298.2	21.7	465	10	US-09-864-761-649	Sequence 849, App
c	44	292.8	21.3	597	10	US-09-864-761-8370	Sequence 8370, Ap
c	45	290	21.1	471	10	US-09-864-761-819	Sequence 819, App

ALIGNMENTS

RESULT 1  
US-09-997-610-1  
: Sequence 1, Application US/09997610  
: Patent No. US20020156244A1  
: GENERAL INFORMATION:  
: APPLICANT: Fox, Brian  
: APPLICANT: Holloway, James L.  
: TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN  
: TITLE OF INVENTION: ZACRP13  
: FILE REFERENCE: 00-96  
: CURRENT APPLICATION NUMBER: US/09/997.610  
: CURRENT FILING DATE: 2001-11-29  
: PRIOR APPLICATION NUMBER: US 60/253,924  
: PRIOR FILING DATE: 2000-11-29  
: NUMBER OF SEQ ID NOS: 7  
: SOFTWARE: FastSeq for Windows Version 3.0  
: SEQ ID NO 1  
: LENGTH: 1381  
: TYPE: DNA  
: ORGANISM: Homo sapiens  
: FEATURE:  
: NAME/KEY: CDS  
: LOCATION: (2)...(1381)  
US-09-997-610-1

Query Match	100.0%;	Score 1377;	DB 9;	Length 1381;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1377;	Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;	
Qy	1	ATAGTGGTCATACCTGTTTAAATACGCCAGTCATTTAGCATGTAGAAAGTTGCTGGACCT	60	
Db	2	ATAGTGGTCATACCTGTTTAAATACGCCAGTCATTTAGCATGTAGAAAGTTGCTGGACCT	61	
Qy	61	CCAGCACACCCCGCCCGCCAGAGAGAGTGGGGCTCTCTGTGGCACCAGGTTTACCACAA	120	
Db	62	CCAGCACACCCCGCCCGCCAGAGAGAGTGGGGCTCTCTGTGGCACCAGGTTTACCACAA	121	
Qy	121	TATACAGGAGAAATAAGTGAATGACAAATGCCCTCTCTCTGATATAGAAAGGTCAGCC	180	
Db	122	TATACAGGAGAAATAAGTGAATGACAAATGCCCTCTCTCTGATATAGAAAGGTCAGCC	181	
Qy	181	TTTACTGTGAAGCTCAGTGGAAAACCTTCTCTTCTTCAAGCCCATCATCTTCACAGG	240	



```
Db 895 TTCAGCTGGCTGCAGAAATTTGCATATGTAATGAGGAGCTGAATGTTAAATCCTCAAGAC 954
Qy 601 AATGGGAAAAATATCTCTGTCGACATGTCCAGAGTCTTCCACAGCAGTCCATCAAAATCACTG 660
Db 955 AATGGGAAAAATATCTCTGTCGACATGTCCAGAGTCTTCCACAGCAGTCCATCAAAATCACTG 1014
Qy 661 GCCTGGAGGCTAGAGAAAATGTTTGTGGGACAGGCCAGGGTCCCTGTGCTGTG 720
Db 1015 GCCTGGAGGCTAGAGAAAATGTTTGTGGGACAGGCCAGGGTCCCTGTGCTGTG 1074
Qy 721 CAGCTAGAGACTTGGTCCCTGTGTCCTCCAGTAAATTCAGCTGTGGCTTCAGAGGGTGCA 780
Db 1075 CAGCTAGAGACTTGGTCCCTGTGTCCTCCAGTAAATTCAGCTGTGGCTTCAGAGGGTGCA 1134
Qy 781 AGCCCCAAGCCTTGGCAGCTTCCCAAGTGGTGTGAGCCTGTGGGTGCAAGAAGTCAAGA 840
Db 1135 AGCCCCAAGCCTTGGCAGCTTCCCAAGTGGTGTGAGCCTGTGGGTGCAAGAAGTCAAGA 1194
Qy 841 ATTGAGGTTTGGGAACCTCCCAATCAGATTTCCAGAGATATATGAAACCCCTGGATGCC 900
Db 1195 ATTGAGGTTTGGGAACCTCCCAATCAGATTTCCAGAGATATATGAAACCCCTGGATGCC 1254
Qy 901 AGGCAGAGTTTGTGTAGGGTGGGTCCCTCATGGGAACCTCTGCAAGGGTACTACAA 960
Db 1255 AGGCAGAGTTTGTGTAGGGTGGGTCCCTCATGGGAACCTCTGCAAGGGTACTACAA 1314
Qy 961 AAGGGAATGTTGGGTGGGAGCCCCACACAGAGTCCCAAGTGGGTCCATCTAGTAGA 1020
Db 1315 AAGGGAATGTTGGGTGGGAGCCCCACACAGAGTCCCAAGTGGGTCCATCTAGTAGA 1374
Qy 1021 GCTGTGAGAAGTCCCAACCTCCAGACTCCAGAGTCCCAAGTGGGTCCATCTAGTAGA 1080
Db 1375 GCTGTGAGAAGTCCCAACCTCCAGACTCCCAAGTGGGTCCATCTAGTAGA 1434
Qy 1081 CAGCATGTGCTGAAATCCACAGACTCAGTCCAGCCTGTGAAGCAGCAGGGATG 1140
Db 1435 CAGCATGTGCTGAAATCCACAGACTCAGTCCAGCCTGTGAAGCAGCAGGGATG 1494
Qy 1141 GAGTCTGTACCTCAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTC 1200
Db 1495 GAGTCTGTACCTCAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTC 1554
Qy 1201 TTGATTTGATGACCTGGAGCTGAGACATGAGATGCAAAAGAGATCATTTTGGAGCTTTA 1260
Db 1555 TTGATTTGATGACCTGGAGCTGAGACATGAGATGCAAAAGAGATCATTTTGGAGCTTTA 1614
Qy 1261 AGATTGACTGCCCTGATTTGCGACTTATATGGGGCCGCTACCCCTTTGTTGGC 1320
Db 1615 AGATTGACTGCCCTGATTTGCGACTTATATGGGGCCGCTACCCCTTTGTTGGC 1674
Qy 1321 CAATTTTTCATTTGGAACCTGCGGTATTTACCCAATGCCCTGACCTCCATTTGTATG 1377
Db 1675 CAATTTTTCATTTGGAACCTGCGGTATTTACCCAATGCCCTGACCTCCATTTGTATG 1731
```

RESULT 3

```
US-09-997-610-3
; Sequence 3, Application US/09997610
; Patent No. US2002015624A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; TITLE OF INVENTION: ZACRP13
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997.610
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
```

```
; LENGTH: 1377
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide encoding a polypeptide
; OTHER INFORMATION: of SEQ ID NO:2
; NAME/KEY: variation
; LOCATION: (1)...(1377)
; OTHER INFORMATION: Each n is independently A, T, G, or C.
; NAME/KEY: misc_feature
; LOCATION: (1)...(1377)
; OTHER INFORMATION: n = A,T,C or G
US-09-997-610-3
```

```
Query Match 71.8%, Score 988.4; DB 9; Length 1377;
Best Local Similarity 58.8%; Pred. No. 3.4e-288;
Matches 810; Conservative 312; Mismatches 255; Indels 0; Gaps 0;

Qy 1 ATAGTGGTCATACCTGCTCTTAATAAGCGGAGTCATTGAGCATGTAGAAGTTGCTGGACCT 60
Db 1 ATHGTNGTNATHCCNGTNTNATHACNGCNGTNTNATHGARCAYGTNGARGTNGCNGNCCN 60
Qy 61 CCAGCACACCCAGGCCCCGAGAGAAGTGGGGCCTCCTGTGGTGCACAGGTTTACCACAA 120
Db 61 CCNGCAYCCNMGCCNCCNGARGGTNGCNGCNGCNGCNGCNGCNGTNTNCNCAR 120
Qy 121 TATACAGGAGAAATAGTGAATGAAATGAAATGCCCTCTCTGTATATAGAAAGTCAAGC 180
Db 121 TAYACNGGARGATHTWSNGARATGACNAARTGYCCNTGYCCNGAYATHCARMGNSGCN 180
Qy 181 TTTACTGTGAAGCTCAGTGGAAACTTCTCTTCTTCAAGCCCATCATCTTCACAGG 240
Db 181 TTYACNGTNAARTYTNWSNGNAARTYTNCCNTTNCNTTAAARCCNATHATHTYACNGN 240
Qy 241 GTCCTGTACAATGCCAGGAGGATTTAAAGGAGGCCATCGGGAGTCTTGTCTGTCAGG 300
Db 241 GTNTNTATAYAGCNCARMNGAYTNAARGAGCNATGGGNGTNTTTCNTGYMNGTN 300
Qy 301 CTGGGAATTACTACTCCAGCTTTGATGTGAGCTGCATCATTTGCAAGTGCAATTTTGG 360
Db 301 CCNGNAAYTAYTAVWSNWSNTTYGAYTNGARYTNCAYCAVTGYAARGTNAAYATHTGG 360
Qy 361 CTAATGAGGAAGCAAAATTTGGCTAATAAGGAAGAAATTTCTAAGCAGCAAGATTCAA 420
Db 361 YTNATGMAARCARATHYTNCAAYAAARGAGARATHWSNAARCARCARWSNATHCAR 420
Qy 421 GAGTGACTCTGGTCTGTTAAAGCATTCAGTCTTCAAGGGAGGAGCAGATGAAGT 480
Db 421 GARGTNACNTGGTNTYTNNAARCCNTYTNWNTTATYHMGNGARGCNGARAYARWSN 480
Qy 481 TCAGAAAATTTGCACCCCTGACAATGTATAAAAAAGAAACCCCATTTTCTGAGGGAAA 540
Db 481 WNGARAAYTNCAYCCNGAYAAAYGTNATHAARAARAARAAYCCNTTWSNGARGNAAR 540
Qy 541 TTCAGCTGGCTGCAGAAATTTGCATATGTAATGAGGAGCTGAATGTTAATCCTCAAGC 600
Db 541 TTYAARTNGCNGCNGARATHGTGATHTGYAAYGARGARYTNAAYCTNAAYCCNCARGAY 600
Qy 601 AATGGGAAAATATCTCTCGGACATGTCCAGAGTCTTCCACAGCAGTCCCATCAATCACTG 660
Db 601 AAYGGNGARAAYATHWSNTGGACNTGYTCARMGNWNSNCARCARWSNATHAARWSNTY 660
Qy 661 GCCTGGAGGCTAGGAGAAAATGTTTGTGGCAGAGCCAGGGTCCCTGTGCTGTG 720
Db 661 GCNTGGMNCNCNMGNNAARTGGTGTGYGNGACNGCNGCNGNWSNTYTGTYGTGN 720
Qy 721 CAGCTTAGAGACTTGGTCCCTGTGTCCTCAGTAAATTCAGCTGTGCTTCAGAGGGTCA 780
Db 721 CARCCNMGNAYTYTNGTNCCTGTGCTNCCNGTNAAYWSNGCNGTNGCWNKNSCARGNGCN 780
Qy 781 AGCCCCAAGCCTTGGCAGCTTCCCAAGTGGTGTGAGCCTCTGGGTGCAAGAAGTCAAGA 840
Db 781 WSNCCNAARCCNTGGCARYTNCNWSNGNGTNGARCCNTNGNGCNGCNAARAARWSNMG 840
```

```
QY 841 ATTGAGTTTGGGAACCTCCAATCAGATTTCAGAAGATATATGAAACCCCTGGATGCC 900
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 841 ATHGARGTNTGGARCCNCCNATHMGNTTYCARARATHAYGGNAAYCCNTGGATGCCN 900

QY 901 AGCGAGAAGTTTGTGTAGGGTGGGGTCTCATGGAGAACCTCTGCAAGGGTAGTACAA 960
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 901 MGNCAARAATTTTCGNTGNGTNGNWSNWSNTGGMGNACNWSNCGNMGNTGNTNCAR 960

QY 961 AAGGGAATGTGGTGGGAGCCCCACACAGAGTCCCCAGTGGGGCTCCATCTAGTAGA 1020
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 961 AARGGNAAYGTNGTGGGARGCCNCCNAYMGNGTNCCNWSNGGNCNCCNWSNWSN 1020

QY 1021 GCTGTGAGAAGAAGTCCACCATCTCCAGACTCCAGAGGGTAGATCCACTCACAGCTTG 1080
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1021 GCNGTNGMGNWNSNCCNCCNWSNWSNMGNTNCAARAGGNMGNWSNACNAYWSNYN 1080

QY 1081 CAGCATGTGCGCTGAATAATCCACAGACACTCAGTCCAGCCCTGTGAAAGCAGCAGGATG 1140
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1081 CARCAYGTNCCNGARAARWSNACNAYACNCARTGCARCCNGTNAARCGNCGNGNATG 1140

QY 1141 GAGTCTGTACCTACAAAACCTAGTGGCAGAGCTGACCAAGACCTGGGAATCTACCTC 1200
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1141 GARWSNGTNCCTAYAARACNCTGNTGNCNGARTNACNAARACNCTNGGNATHAYTYN 1200

QY 1201 TTGCATTTGTCATGACCTGGACCTGAGACATGAGTCAAAAGAGATCATTTTGGAGCTTTA 1260
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1201 YTNCAYTGYCAGYTNGAYTNGMNCAYGGNGINGNABMGNGAYCAYTTYGNGCNYN 1260

QY 1261 AGATTGACTGCCACCTCGGATTTCCGAGCTTATATGGGGCCCGTACCCCTTTGTTGGC 1320
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1261 MGNTTYGAYTGYCCNACNGNTTYMGNACNTAYATGGGNCNCTNCCNTNTGYTTYGN 1320

QY 1321 CAATTTTTCATTTGGAACTGCCGTATTTACCAATGCTCTACTCCATGATGATG 1377
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1321 CARTTTTCCNTTYGGNACNCGNTNTTYACNCARTGIYNTAYTNCAYTGYATG 1377

RESULT 4
US-09-997-610-7
; Sequence 7, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; TITLE OF INVENTION: ZACRPL3
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1731
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide sequence of zacrp13/zhpl
; OTHER INFORMATION: of SEQ ID NO:6
; NAME/KEY: misc_feature
; LOCATION: (1)...(1731)
; OTHER INFORMATION: n = A,T,C or G
US-09-997-610-7
```

Query Match 71.8%; Score 988.4; DB 9; Length 1731;  
Best Local Similarity 58.8%; Pred. No. 3.9e-286;  
Matches 810; Conservative 312; Mismatches 255; Indels 0; Gaps 0;

```
QY 1 ATAGTGGTCATACCTGCTCTTAATACGGCAGTCATTTGACGATGTAGAAAGTTGCTGGACCT 60
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 355 ATHGTNGTNAHCCNGNTNTTNAHCCNGCNGTNGATHGARCAYGTNGRTNGCNGNCCN 414
```

```
QY 61 CCAGCACACCCAGGCCCCGAGAGAAAGTGGGCTCTCTGGTGCACAGGTTTACCACAA 120
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 415 CCNGCNCACTCCNGNCCNCCNGARGTNGNCCNCCNGNCCNGNCCNGNTNCCNCAR 474

QY 121 TATACAGGAGAATTAAGTGAATGACAAAATCCCTGCTGATATAGAAAGGTCAGCC 180
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 475 TAYACNGGNGARATHWSNCGARTGACNAARTGCCNTGYCCNGAYATHGARMGNWSGNCN 534

QY 181 TTTACTGTGAAGCTCAGTGGGAAAACCTTCTCTCTTCAAGCCCATCATCTTCACAGC 240
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 535 TTYACNGTNAARTYTNWSNGNNAARTNCCNTTYAACRCCNATHATHTTYACNGN 594

QY 241 GTCTGTGTACAAATGCCAGAGGATTTAAAGGAGGCATGGAGTCTTGTCTTTCGAGGGTG 300
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 595 GTNYNTATAYATGCNARMNGNAYYTNAAARGAGCNGATGGNGTNTTYCNGTYMGNGTN 654

QY 301 CCTGGGAATTTACTCTCCAGCTTTGATGTTGAGCTGCATCATTTGCAAGGTGAATATTTGG 360
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 655 CCNGGNAAYTAVTAYWSNWSNTTYGAYGTNGARTNCAVCAVTGYAARTNAAATHTGG 714

QY 361 CTAATGAGGAAGCAAAATTTTGGCTAATAAGGAAGAAATTTCTAAGCAGCAAAAGCATCAA 420
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 715 YTNATGMGNARCARATHYTNCGNCAAYAAARGAGARATHWSNAAARCARCARNSNATHCAR 774

QY 421 GAGGTGACTTGGTGTCTGTTAAAGGCATTCAGTTTCATAAGGGAGGCAGAGCATAAAGCT 480
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 775 GARGTNACNTGGGTNTYTNNAARGCNTTYWSNTTYATHWNGARGCNGARCAAYAARWSN 834

QY 481 TCAGAAAATTTCCACCTGCACAAATGTGATAAAGAAAGAAACCCATTTTCTGAGGGGAAA 540
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 835 WSGNARAAYTNCAYCCNGAYAAAYGTNATHAARAARAARAYCCNTTYWSNGARGNAAAR 894

QY 541 TTTCAAGCTGGCTGCAGAAATTTGCATATCTAATGAGGAGCTGAATGTTAATCTCAAGAC 600
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 895 TTYAARYTNGCNGCNGARATHGYATHGTGYAAGARGARYTNAAYGTNAAAYCCNCARGAY 954

QY 601 AATGGGAAAATATCTCTGGACATGTCAGAGGTCTTCACAGAGTCTTCACAGGTCCTCAAACTCAGTG 660
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 955 AAYGGNGARAAYATHWSNTGGACNTGYCARMGNWSNWSNCAARCAARWSNATHAARWSNYN 1014

QY 661 GCCTGGAGGCTTAGGAGAAAATGGTTTGTGGGACAGGCCAGGGTCCCTGTGCTGTGG 720
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1015 GCNTGGMGNCCNMGMNNAARTGGTTTYTGYNACNGNCCNGNWSNNTNTGYTYGYN 1074

QY 721 CAGCTTAGAGACTTGGTGGCTGTCTCCAGTAAATTCAGCTGTGCTTCAGAGGTGCA 780
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1075 CARCCNMGNAYTNGTNCCTGYTNCNGTNAAYWSNCGNCTGNCNWSNGARGNGCN 1134

QY 781 AGCCCCAAGCCTTGGCAGCTTCCAAGTGTGTGAGCCTGTGGGTGCAAGAGAGTCAAGA 840
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1135 WSNCCNAARCCNTGGCARYTNCCNWSNGNGINGARCCNTNGGNCNNAARAARWSNMG 1194

QY 841 ATTGAGTTTGGGAACCTCCAATCAGATTTCAGAGATATATGGAACCCCTGGATGCC 900
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1195 ATHGARGTNTGGGARGCCNCCNATHMGNTTYCARAARATHAYGGNAAAYCCNTGGATGCCN 1254

QY 901 AGSCAGAAGTTTGTCTAGGGTGGGTCTCATGAGAACCTCTCAGAGGTGTAGTACAA 960
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1255 MGNCAARAATTTGCGTNGTNGNNGNWSNWTGGMGNACNWSNCGNMGNTGNTNCAR 1314

QY 961 AAGGGAATTTGGGTGGGAGCCCCACACAGAGTCCCCAGTGGGCTCCATCTAGTAGA 1020
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1315 AARGGNAAYGTNGTGGGARGCCNCCNAYMGNGTNCCNWSNNGNCCNCCNWSNWSN 1374

QY 1021 GCTGTGAGAAGAAGTCCACCATCTCCAGACTCCAGAGGGTAGATCCACTGCACACTTG 1080
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1375 GNGTNGMGNWSNCCNCCNWSNWSNMGNTNCAARAARGNMGNWSNACNAYWSNYN 1434

QY 1081 CAGCATGCTCCGTAAGAAATCCACAGACACTCAGTGGCAGCCTGTGAAAGCAGCAGGATG 1140
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1435 CARCAYGTNCCNGARAARWSNACNNGAYACNCARTGYCARCCNGTNAARCGNCGNCGNATG 1494
```

```
QY 1141 GAGTCTGTACCTACAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTC 1200
Db 1495 GARNSGTNCCNTAAYAAACNGTNGTNGCNGARYTNACNAARACNGTNGGNATHAYTNTN 1554
QY 1201 TTGATTTGTATACCTGAGCTGAGACATGAGTCAAAAGAGATCATTTTGGAGCTTTA 1260
Db 1555 YTCAYTGCAAYAYNGAYGTWNGCAYGNGTNAARMNGNGAYCATTTTGGNGCNTN 1614
QY 1261 AGATTTGACCTGCCACTGGATTTTGGACATTTATATGGGGCCCGTACCCCTTTTGGG 1320
Db 1615 MNTTYGAYTGCCNACNGGNTTGMGNACNTAYATGGGNCNGTNCNTNTGTGTYGNN 1674
QY 1321 CAATTTTTCATTTGGAACCTGCGTATTTACCCCAATGCCCTGACCTCCATGTATG 1377
Db 1675 CARTTYYCCNTTYGGNACNGCNGTNTTYACNCARTGYTNTAYTNCAYTGYATG 1731

RESULT 5
US-10-026-188-6/c
: Sequence 6, Application US/10026188
: Patent No. US2002016465A1
: GENERAL INFORMATION:
: APPLICANT: Zhang, Yifeng
: APPLICANT: The Regents of the University of California
: TITLE OF INVENTION: Assays for Taste Receptor Cell Specific
: TITLE OF INVENTION: Ion Channel
: FILE REFERENCE: 02307E-114910US
: CURRENT APPLICATION NUMBER: US/10/026,188
: CURRENT FILING DATE: 2001-12-21
: PRIOR APPLICATION NUMBER: US 60/259,379
: PRIOR FILING DATE: 2000-12-29
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 6
: LENGTH: 155074
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: Human genomic region containing ltrpc5
: OTHER INFORMATION: (Human Chromosome 11p15.5 PAC clone pD915f1
: OTHER INFORMATION: containing KvLQT1 gene)
US-10-026-188-6

Query Match 46.6%; Score 641.6; DB 9; Length 155074;
Best Local Similarity 80.7%; Pred. No. 3e-182;
Matches 826; Conservative 0; Mismatches 159; Indels 39; Gaps 5;

QY 390 GGAAGAAATTTCTAAGCAGCAAAAGCATTCAGAGGTGACTTGGGTGCTGTTAAAGGCATT 449
Db 94901 GGAAGAAATTTCTAAGCTGCAAGCTTCAAGAGGTGACTTGGGTGCTGTTAAAGGCATT 94842
QY 450 CAGTTTCATTAAGGAGCAGCAGCATAGATTTCAGAAAATTTGACCTGCACATGTGAT 509
Db 94841 CAGTTTTCAGAGGGAACAGAG-ATAAAGTTCAGAAAATTTGACCTGCACATGTGAT 94783
QY 510 AAAAAAGAAAACCCATTTCTGAGGGGAAATTCACGCTGCTCAGAAAATTTCCATATG 569
Db 94782 AGAAAATAAAACCCATTTCTGAGGAGAAAGTCAACGACCTGCAGAAAATTTCCATTAAG 94723
QY 570 TAATGAGGAGCTGAATTTAATCCTCAAGACAATGGGGAATAATCTCCTGGACATGTCA 629
Db 94722 TAACAAGAACCAATGTTAATCACTAAGACAATGGGGAATAATCTCCTAGGCGCATGTCA 94663
QY 630 GAGCTTTCACAGCAGTCCATCAATACATCGGCTGGAGGCGCTAGAG-AAAATGGTT 686
Db 94662 GAGACCTTAGGAGCAGGCGCTCCCATCCAGGCCAGAGGCGCTAGGAGGCAAAAATGGTT 94603
QY 687 TTGTGGGAGCGCCAGGGTCCCTGTGCTGTGTGCAGCTAGACATTTGGTCCCTGTGT 746
Db 94602 TCCTGGACTGGGCCAGGGCTCCATCTGTGTGTGCAGCTAGGAGCTTGGTGGCCCTGTGT 94543
QY 747 CCCAG-----TTAATTCAGCTGTGGTTCAGA 773
```

```
Db 94542 CCCAGCCGCTCTAGCCATAGCTAAAAAGGTGTCAAGGTACAGCTTGGGCTGTGGCTTCAGA 94483
QY 774 GGGTGAAGCCCCCAAGCTTTGGCAGCTTTCCAAAGTGGTGTGAGCCTGTGGTGAAGAA 833
Db 94482 GGGTGAAGCCCCCAAGCTTTGGTAGCTTCCATGTGGTGTGAGCCTGTGGTGAAGAA 94423
QY 834 GTCAGAAATTGAGGTTTGGGAACCTTCCAAATCAGATTTTCAGAAAGATATATGAAACCCCTG 893
Db 94422 GTCAGAAATTGAGGTTATGGAAGCTCTGCCCTAGATTTTCAG-AGATGTATGGAATGCTGTG 94364
QY 894 GATGCCAGGAGCAAGTTTGTCTAGGGGTGGGTGCTTCATCGGAGAACTCTGCAAGGCT 953
Db 94363 GATGTCCAGGAGCAAGTTTGTCTAGGGGTGGGTGCTTCATCGGAGAACTCTGCTAGGCT 94304
QY 954 AGTACAAAAGGAAATGTTGGTGGGAGCCGCCACACAGAGTCCCCAGTGGGGCTCCATC 1013
Db 94303 AGTGAGGAGGAGAAAGTGGGGTGGAGCCGCCATGCAAAAGTCCCACTGGGGCACTGCC 94244
QY 1014 TAGTAGAGCTGTGAGAAAGTCCACCACTCTCCAGACTCCAGAGAGGGTAGATCCACTGA 1073
Db 94243 TAGTGGAGCTGTGAGAGAGAGGCCACCTGCTCTCCAGATCCCAAGATGATAGATCCATCA 94184
QY 1074 CAGCTTGCAGCATGTGCTGAAATTCACACAGACACTCAGTGCAGCCTGTGAAAGCAGC 1133
Db 94183 CAGCTTGCAGCATGTGCTGAAAGTGCAGACACTCAATGCCAGCCTGTGAAAGCAGC 94124
QY 1134 AGGATGGAGTCTGTACCTACAAAACCGTAGTGGCAGCTGACCAAGACCGTGGGAAT 1193
Db 94123 CAGGAGGAGGCTGTACCTGCAAAAGCAGAGGGGAGAGCTGCTCAAGACTATGGGAAC 94064
QY 1194 CTACTCTTTCATGCTCATGACCTGGAGCTGAGACATGGAGTCAAAAGAGATCATTTTGG 1253
Db 94063 CCACCTCTGTCATCAGCATGACCTAGATGTGAGACATGAGCTCAATGGAGATCATTTTGG 94004
QY 1254 AGCTTTAAGATTTGACTGCCCTCCACTGGATTTTCGGACTTATATGGGGCCCGTA-CCCCTTT 1312
Db 94003 AGCATTAAGATTTAACTGCCCTCCACTGGATTTTCGAACCTTGCATGGGGCCCTTTAGCCCTTC 93944
QY 1313 GTTTTGGCAATTTTTCATTTGGAACCTGCGGTATTTTACCAATGCTGTACCTCCATT 1372
Db 93943 GTTTTGGCAATTTTACCTCCATTTCAAAATGGGTGTTATTTTCAAGGCTGTACCTCAATT 93884
QY 1373 GTAT 1376
Db 93883 GTGT 93880

RESULT 6
US-09-933-267A-1
: Sequence 1, Application US/09933267A
: Patent No. US20020123095A1
: GENERAL INFORMATION:
: APPLICANT: Kalush, Francis et al.
: TITLE OF INVENTION: Estrogen receptor alpha variants and
: TITLE OF INVENTION: methods of detection thereof
: FILE REFERENCE: CL000258C14
: CURRENT APPLICATION NUMBER: US/09/933,267A
: CURRENT FILING DATE: 2001-08-21
: PRIOR APPLICATION NUMBER: 60/160626
: PRIOR FILING DATE: 1999-10-20
: PRIOR APPLICATION NUMBER: 60/183756
: PRIOR FILING DATE: 2000-02-22
: PRIOR APPLICATION NUMBER: 09/692414
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 09/768184
: PRIOR FILING DATE: 2001-01-24
: PRIOR APPLICATION NUMBER: 09/804076
: PRIOR FILING DATE: 2001-03-13
: PRIOR APPLICATION NUMBER: 09/826314
: PRIOR FILING DATE: 2001-04-05
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 1
; LENGTH: 465237
; TYPE: DNA
; ORGANISM: human
US-09-933-267A-1

Query Match      44.9%; Score 618.2; DB 10; Length 465237;
Best Local Similarity 78.2%; Pred. No. 6.6e-175;
Matches 847; Conservative 0; Mismatches 193; Indels 43; Gaps 7;

QY 334 CTGCATCATTCGAAGGTGAATATTTGGCTAATGAGGAACAAATTTTGGCTAATAAGGAA 393
Db 438730 CTGCTGTAGAGATTTGTGGAATTTGAACCTTGAGAGAAATGATTTGGGTACCTGCTAA 438789

QY 394 GAAATTTCTAAAGCAGCAAGCAATCAAGAGGTGACTTGGGTGCTGTAAAGGCATTCAGT 453
Db 438790 GAAATTTCTAAAGCAGCAAGCAATCAAGAGGTGACTTGGGTGCTGTAAAGGCATTCAGT 438849

QY 454 TTCTAATAGGGAGGCGAGCATAAAGAGTTTCAGAAAAATTTGCACCCCTGCAGAAATGTGATAAA 513
Db 438850 TT--TAAAGAGCAAAACAGCATAAAAGTTTCAAGAAATTTGCAGCCCTGATGATCGAGTAGGA 438907

QY 514 AGAAAAACCCCA-TTTTCTGAGGGGAAATTCAGCTGCGCTGCAGAAATTTGCATATGTAA 572
Db 438908 AGAAAAACCCCAATTTTCTGAGGAGAAATTCAGCTGCGCTGCAGAAATTTGCATATGTAA 438967

QY 573 TGAGGAGCTGAATGTTAATCCTCAAGACAATGGGAAAAATATCTCCTGGACATGTCAGAG 632
Db 438968 CAAGGAGCAATGTTAATCCCAAGACAAATGGGAAAAATGTTCTCCAGAGCATGTCATAG 439027

QY 633 GTCTTTCACAGAGTCCCATCAATCACTGGCCCTGGAGGCCCTAGGAG-----AAAATGGTTTT 688
Db 439028 GTCTTTCATGCGACGCCCTCCCATCACAGACCGCGGAAGCCTAGGAGGAAAAAACAGTTTT 439087

QY 689 GTGGACAGGCCCCAGGTCCTCTGCTGTGTCAGCCTAGAGACTTGTGGCCCTGTGTGCC 748
Db 439088 GTGGGCCCATGCCCCAGGGTCCCATGCTGTGTGACGCTAGGAACTTGTGGCCCTGCATCT 439147

QY 749 CAGTTAATTCAGCTGT-----GCTTCAGAGGGGTGCAAGCC 784
Db 439148 CAGCTGTCTCCAGCTATTTGCTAAAGGGGCTGAGGTACCCAGGTTTTCAGAGTTTGCAGGCC 439207

QY 785 CCAAGCCTTTGGCAGCTTCCAAGTGGTGTGTAGCCCTGTGGTGCAAGAGAGTCAAGAAATG 844
Db 439208 CCAAACTTTGGCAGCTTTCATCTGTGTGTGTAGCCCTGTGTGTACAGAGAGTTTGAAGATTTG 439267

QY 845 AGGTTTGGGAACCTCCAATCAGATTTCAAGAGATATATAGGAACCCCTGGATGCCCAGGC 904
Db 439268 AGGTTTGGGAACCTCCAATCAGATATATTTCAAGAGATACGTGGAAATGCCCTGGATACCCAGGC 439327

QY 905 AGAAGTTTGTCTAGGGGTGGGTGCTCATGAGAACCTCTCAAGAGGTAGTACAAAAGG 964
Db 439328 AAACATTTGTCTCAGAGTTGGGGCCCTCATGAGGGCCCTCTGCTAGGCAATGAGGAAGG 439387

QY 965 GAAATTTGGGTGGAGGCCCCACACAGAGTCCCAGTGGGCTCCATCTAGTAGAGCTG 1024
Db 439388 GAAATTTGGGGTTGGAACCCCCACACAGAGTCCCCACTGGGCACTGCCCTAGTGGAGCTG 439447

QY 1025 TCAGAGAAGTCCACCATCTCCAGACTCCAGAGGGTAGATCCACTGACAGCTTGCAGC 1084
Db 439448 TAAGAGGAGGACCACTGTCTCTCCAGACCGCAGAAATAGTAGATCCACTGACAGCTTGCAC 439507

QY 1085 ATGTGCTGAAAAATCCACAGACACTCAGTGCACGCTGTGAAGCA-----GCA 1134
Db 439508 ATGTGCTGGAANAAGCCACAGACACTCAACGCCAGCCCTGTGAAGCAAGTTCAGGGTTGGAG 439567

QY 1135 GGGATGGAGTCTGTACCCCTACAAAAACCCCTAGTGGCAGAGCTGACCAAGACCCGTGGGAATC 1194
Db 439568 GTGGTGGTGGCTATACCCCTATAAGCCACAGGCGCAGAGCTGCCCAAGACTATGGGAACC 439627

QY 1195 TACCTCTTGCATTTGCATGACCTGAGCGTGTGAGACATGAGGAGTCAAAAGAGATCATTTTGA 1254
Db 439628 TACCTCTTGCATCAGCATGACCTGGATGTGTGAGACATTTAGTCAAAAGGAGAT-ATTTTGA 439686
```

```
QY 1255 GCTTTAAGATTTGACTGCCCACTGGATTTGCGACTTATATGGGGCCCGT-ACCCCTTTG 1313
Db 439687 GCTTTAAGATTTGACTGCCCTGGTGGATTTAGACTTGTGTGGCCCTGTAAACCCCTTTG 439746

QY 1314 TTTTGGCCAAATTTTTCATTTGGAACCTGCCGTATTTACCCAATGCCTGTACTCCATTTG 1373
Db 439747 TTTTGGCCAAATTTCTCCCAATTTGGAGCTGCTGTATTTACCCAATGCCTAAACCCGCAATG 439806

QY 1374 TAT 1376
Db 439807 TAT 439809

RESULT 7
US-10-003-806-6
; Sequence 6, Application US/10003806
; Patent No. US2002011929A1
; GENERAL INFORMATION:
; APPLICANT: Bishop, Colin E.
; APPLICANT: Agoulnik, Alexander I.
; APPLICANT: Zhu, Qichao
; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
; FILE REFERENCE: P020660SI/10024824
; CURRENT APPLICATION NUMBER: US/10/003,806
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 60/245,872
; PRIOR FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 6
; LENGTH: 180557
; TYPE: DNA
; ORGANISM: Human
US-10-003-806-6
```

```
Query Match      44.3%; Score 610.2; DB 12; Length 180557;
Best Local Similarity 78.9%; Pred. No. 9.8e-173;
Matches 809; Conservative 0; Mismatches 168; Indels 48; Gaps 5;

QY 390 GGAAGAAATTTCTAGCAGCAAGCAATTCAGAGGTGACTTGGGTGCTGTTAAAGGCATTT 449
Db 122662 GGAAGAAATTTCTAAGCAGCAAAACATTTGAAGAGGTGACTTGGGTGCTATTTAAAGCATT 122721

QY 450 CAGTTTTCATAAGGGA---GGCAGAGCATAAAGATTTTCAGAAAAATTTGCACCCCTGACAATCT 506
Db 122722 TAGTTTTAAAGGGAAACAGCATAAATTAAGTTTCAAAAAATTTGCACCCCTGACATGCT 122781

QY 507 GATAAAAAAGAAAAACCCCATTTTCTGAGGGGAAATTCAGAGCTGGCTGCAGAAATTTGCAAT 566
Db 122782 GATAGAAAAAGAAATCCCGTTTTTCTGAGCAGAAATTCAGGCCAGCTGCAGAAATTTGCAAT 122841

QY 567 ATGTAATGAGGAGCTGAATGTTAATCCTCAAGACAATGGGAAAAATATCTCTCGNACATG 626
Db 122842 AAGTAACAGAGGACCAAAATGTTAGTCCCAAGATAATGGGAAAAATGTTCTCCAGGGCATG 122901

QY 627 TCAGAGGTCTTCACAGCACTCCATCAAAATCACTGGCTGGAGGCCCTAGCAGAAAAATGGTT 686
Db 122902 CCAGAGGTCTTCACGCGACCCCACTCACATAACAGCCCTGGAGGTGTAGGAGGAAAAAAGT- 122960

QY 687 TTTGGGACAGGCCCCAGGGTCCCTGTGCTGTGTGCAGCCCTAGAGACTTGGTCCCTGTGT 746
Db 122961 -----GGGCACAGGGTCCCGACACTGTGTGCAGGCTAGGACTTGGTCCCTGCAT 123011

QY 747 CCCAGTTAATTCAGCT-----GTGGCTTCAGA 773
Db 123012 CCCAGTGTCTCCAGCTGTGGCTGAAAAAGGCCAGTGTAGAGCTTGGGCCCTGTGGTGCAGA 123071

QY 774 GGGTCAAGCCCCAAGCCCTTGCAGCTTCCAAGTGGTGTGAGCCCTGTGGTGCAGAAAG 833
Db 123072 GGGTCAAAACCCCAAGCCCTTGGCAGCTTCCACATTTGTTGAGCCCTGTGGCAGTGCAGAA 123131

QY 834 GTCAAGAATTTAGGTTTGGGAACCTCCCAATCAGATTTTCAGAAGATATATATGAAACCCCTG 893
```

```

Db 123132 GTCCAGAACCTGAGGTTTCAGAACTCGCGCTAGATTTCAGAGATGATATGGAAATTCCTG 123191
Qy 894 GATCCCGAGGACAGAAATTTGCTGTAGGGGTGGGGTCTCATGAGAGAACCTCTGCAAGGCT 953
Db 123192 GATCCCGAGGACAGAAATTTGCTGCAGGGGTGGGGCTCATGAGAGAACCTCTGTAGGAC 123251
Qy 954 AGTCAAAAGGGAATTTGGGTGGGAGCCCCACACAGAGTCCCCAGTGGGGCTCCATC 1013
Db 123252 AATCGGAAGGGAAGTGTGGATTGGAGTCCCCACACAGAGTCCCTACTGAGGCACCGC 123311
Qy 1014 TAGTAGAGCTGTGAGAGAAGTCCACCATCTCCAGACTCCAGAAGGGTAGATCCACTGA 1073
Db 123312 TAGTGTAGCTGTGAGAAGAGGGCCCATCTTCAGACCCCAAGATGAGTCCACTGA 123371
Qy 1074 CAGCTTCGACATGTGCTGAAAAATCCACAGACACTCAGTGCACGCTGTGAAAGCAGC 1133
Db 123372 CAGCTTCGACATGTGCTGGAAGGCCACAGGCACTCAACACCGCCATGAACCTCAGC 123431
Qy 1134 AGGGATGGAGT-CTGTACCCTCAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAA 1192
Db 123432 CGGGTGGGGGTGCCATACCCTGCANAATCACAGGGTCTAGCTCCCCAGGTCATGGGA 123491
Qy 1193 TCTACCTCTTGCATGTATGACCTGGACGTGAGACATGGAGTCAAAAGAGATCATTTTG 1252
Db 123492 CTCAACTCTTGTATCAGCATGATCTGGATGAGACATGGAGTCAAAAGAGATCATTTTG 123551
Qy 1253 GAGCTTTAAAGATTTGAGTGGCCACATGGATTTGGGACTTATATGGGGCCCGTA-CCCCTT 1311
Db 123552 GAGCTTTAAAGATTTGATTTGGCCCTGCTGGATTTTGGACTTGCATGGGGCTGTAGCTCCCT 123611
Qy 1312 TGTTTTGGCCAAATTTTTCATTGGAACTGCGGTATTTACCAATGCGCTGTACCTCCAT 1371
Db 123612 TGTTTTGGCCAAATTCCTCCCATTTGGAATGGCTGTATTTATCGAATGCCATACCCCTG 123671
Qy 1372 TGTAT 1376
Db 123672 TGTAT 123676

RESULT 8
US-10-003-806-9
; Sequence 9, Application US/10003806
; Patent No. US20020119929A1
; GENERAL INFORMATION:
; APPLICANT: Bishop, Colin E.
; APPLICANT: Agoulnik, Alexander I.
; APPLICANT: Zhu, Qichao
; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
; FILE REFERENCE: P02066051/10024824
; CURRENT APPLICATION NUMBER: US/10/003,806
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 60/245,872
; PRIOR FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 180557
; TYPE: DNA
; ORGANISM: Human
US-10-003-806-9

```

Qy	507	GATAAAAAGAAAAACCCATTTTCTTGAGGGAAATTAAGAGCTGGCTGCAGAAATTTGCAAT	566
Db	122782	GATAGAAAAGAAAATCCGTTTTTCTGAGCAGAAATTAAGGCCAGCTGCAGAAATTTGCAAT	122841
Qy	567	ATGTAANTGAGGAGCTGAATGTTAATCCTCAAGACAAATGGGAAAAATATCTCCTGGACATG	626
Db	122842	AAGTAACAAGGAGCAAAATGTTAGTCCCAAGATAAATGGGAAAAATGTTCTCCAGGCGATG	122901
Qy	627	TCAGAGGTCTTTTCAGCAGCACTCCATCAAAATCACCTGGCTGGAGCCCTAGGAGAAAATGGTT	686
Db	122902	CCAGAGGTCTTCAGGCGAGCCACTCACATAACAGCCCTGGAGGTGTAGGAGGAAAAAGT	122960
Qy	687	TTGTGGACAGGCCAGGGTCCCTGTGCTGTGTGCAGCCCTAGAGACTTGGTGCCCTGTGT	746
Db	122961	-----GGGCACAGGGTCCAGCACTGTGTGCAGGCTAGGCACTTGGTTCCTGCAAT	123011
Qy	747	CCAGTTAATTCAGCT-----TGCGCTTCAGA	773
Db	123012	CCAGCTGTCTCCAGCTGTGGCTGAAAGAGGCCAGTGTAGAGCTTGGGCCGTGGCTGCAGA	123071
Qy	774	GGGTGCAAGCCCCAAGCCTTTGGCAGCTTCCAAGTGGTGTGTGAGCCCTGTGGGTGCAAGAA	833
Db	123072	GGGTGCAAAACCCCAAGCCTTTGGCAGCTTCCAACATGTTGTGTGAGCCCTGTGGGTGCACAGAA	123131
Qy	834	GTCAAGAAATTCAGGTTTGGGAACCTCCAATCAGATTTCAAGAGATATATGGAAACCCCTG	893
Db	123132	GTCCAGAACTGAGGTTTTCAGAACTCCGCCCTAGATTTTCAGAACATGATATGGAAATTCCTG	123191
Qy	894	GATCCCCAGGCAGAAAGTTTCTGTAGGGGTGGGGTCTCATGGAGAACCTCTGCAAGGTT	953
Db	123192	GATCCCCAGGCAGAAAGTTTCTGCAGGGGTGGGGCCCTCATGGAGAACCTCTGTTAGGAC	123251
Qy	954	AGTCAAAAGGGAAATGTTGGGTGGAGGCCACACAGAGTCCCCAGTGGGGCTCCATC	1013
Db	123252	AATCCGAGAGGGAAGTGTGGGATTTGGATCCCCACAGAGTCCCTACTGAGGCACCGCC	123311
Qy	1014	TAGTAGAGCTGTGAGAGAAAGTCCACCAATCCTCCAGACTCCAGAGGGTAGATCCACTGA	1073
Db	123312	TAGTGTAGCTGTGAGAGAGGGCCACCATTCTCCAGACCCCAAGATGGTAGATCCACTGA	123371
Qy	1074	CAGCTTCAGCATGTGCTGAAATAATCCACAGACACTCAGTCCAGCCCTGTGAAGCAGC	1133
Db	123372	CAGCTTCAGCATGTGCTGAAAGAGCCACAGGCACTCAACACGACCCCATGAATCAGC	123431
Qy	1134	AGGGATGGAGT-CTGTACCCTACAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGAA	1192
Db	123432	CGGGTGGGGTGCCATACCTCGCAAAATCACAGGGTCTAGCTCCCCAGGATGCGGAA	123491
Qy	1193	TCTACCTCTTGCAATGTGATGACCTGGACGTGAGACATGGAGTCAAAAGAGATCATTTTG	1252
Db	123492	CTCAACTCTTGATTCAGCATGATCTGGATGTAGACATGGAGTCAAGGAGATCATTTTG	123551
Qy	1253	GAGCTTTAAGATTTGACTGCCCCNACTGGATTTCCGACTTATATGGGGCCGTA-CCCTTT	1311
Db	123552	GAGCTTTAAGATTTGATTTGCCCTGCTGGATTTTGGACTTGCATGGGGCCCTGTAGCTCCTT	123611
Qy	1312	TGTTTTGGCCAAATTTTTCATTTTGAACCTGCGGATATTTACCCAAATGCTGTACCTCCAT	1371
Db	123612	TGTTTTGGCCAAATTCCTCCCAATTTGGAATGGCTGATTTATTCGAATGCCTATACCCCTGT	123671
Qy	1372	TGTAAT	1376
Db	123672	TGTAT	123676

RESULT 9  
US-09-835-232-77C  
; Sequence 7, Application US/09835232  
; Patent No. US20020098489A1  
; GENERAL INFORMATION:  
; APPLICANT: Leder, Philip  
; APPLICANT: Leder, Benjamin

; TITLE OF INVENTION: FORMIN-2 NUCLEIC ACIDS AND POLYPEPTIDES
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 00383/052002
; CURRENT APPLICATION NUMBER: US/09/835,232
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,811
; PRIOR FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 170834
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc\_feature
; LOCATION: (1)...(170834)
; OTHER INFORMATION: n = A,T,C, or G
US-09-835-232-7

Query Match 43.8%; Score 603.4; DB 10; Length 170834;
Best Local Similarity 79.6%; Pred. No. 1.le-170;
Matches 816; Conservative 0; Mismatches 166; Indels 43; Gaps 7;

QY 391 GAAGAAATTTCTAAGCAGCAAAAGCATTCAGAGGTGACTTGGGTGCTGTAAAGGCATTC 450
|||||
Db 113857 GAAGAAATTTCTAAGCAGCAAAAGCATTCAGAGGTGACTTGGGTGCTGTAAAGGCATTC 113798
|||||

QY 451 AGTTTCATAGAGGAGCGACGATAGAGTTTCAGAAAAATTTCCAGCCCTGCACAAATGTGATA 510
|||||
Db 113797 AGTTTAAAGGAAATAGAGCATAAATTTTGGAAAAATTTCCAGCCTGCACAAATGTGATA 113738
|||||

QY 511 AAAAAAGAAAAACCCATTTTCTGAGGGGAAATTCAGAGCTGGCTGCAGAAATTTGCATATGT 570
|||||
Db 113737 GAAAGAAAAAGCTCATTTTCTCAGGAGAAATTCAGAGCGCTGCAGGAATTTGCATAAGT 113678
|||||

QY 571 AATGAGGAGCTGAATGTTAATCCTCAAGACAATGGGAAAAATATCTCTGGACATGTGAC 630
|||||
Db 113677 AATGAGGAGTTGAATGTTAATCACCAGACAATGGGGTAAAGCTCTCCAGGGCATGTAC 113618
|||||

QY 631 AGGTCTTCACACAGTCCATCAATCACTGGCCTGGAGGCCT--AGGAGAAATGGTTT 687
|||||
Db 113617 AGACCTTGGCAGCAGACCTTCCATCAGAGCCGACAGCTTGGGAGGATAAAAATGGTTT 113558
|||||

QY 688 TGTGGACAGGCCAGGGTCCCTGTGTGTGTCAGCCCTAGAGACTTGGTGCCTGTGTC 747
|||||
Db 113557 TGTAGGAGGCCAGGGTCCCTGTGTGTGTCAGCCCTAGGACTTGGTGTCTGTGCATC 113498
|||||

QY 748 C-----CAGTTAATTCAGCTGTGGCTTCAGAGG 775
|
Db 113497 CAGCTGCTCTAGCCATAGCTAAAGGGTCCAAAGGTATAAATCGGCTGTGGCTCTAGAAG 113438
|||||

QY 776 GTGCAAGCCCAAGCCTTGGCAGCTTCCAAGTGGTTTGAGCCTGTGGGTGCAAAAGT 835
|||||
Db 113437 GTGCAAGTCCCAAGTCTTGGCAGCTTCCACGTGATGTGTGAGCTTCCAGGCCACAGAAG- 113379
|||||

QY 836 CAAGAATTGAGGTTTGGGAACCTCCAATCAGATTTTCAGAAAGATATATGGAACCCCTGGA 895
|||||
Db 113378 -AAGAATTGAGGTTTGGGAACCTCCGCCTAGATTTTCAGAGGATATATGGAATGCCTGGA 113320
|||||

QY 896 TCCCAAGCAGCAAGTTTGTGTAGGGGTGGGTCTCATATGGAGAACCTCTGCAAGGGTAT 955
|||||
Db 113319 TCTCTAGGAAGAAGTTTGGCAGAGGGGAGGGTCTCATATAGAAACCTCTGTAGGGCAG 113260
|||||

QY 956 TACAAAAGGGAATGTTGGGTGGGAGCCCCACACAGAGTCCCCAGTGGGGCTCCATCTA 1015
|||||
Db 113259 TCCAGAGGGAATATGGGTTTGGAGCCCCACACAGAGTCCCTACTGGGGCAGTGCCTTA 113200
|||||

QY 1016 GTAGAGCTGTGAGAGAAGTCCACCATCTCCAGACTCCAGAGGGGTAGATCCACTGACA 1075
|||||
Db 113199 GTGAGCTGTGAGAGAAGAGGCCACTGTCTCCAG-CCCAGAAATAGTAGATCCACTGATA 113141
|||||

QY 1076 GCTTGGACAGATGTGCTGAAAAATCCACAGACACTCAGTGGCAGGCTGTGAAAGCAGCAG 1135
|||||

Db 113140 GCTTGACCATGCACCTGGAAAAGCCATAGACACTCAATGCCAGCCCATGAAAGCAGCTG 113081
QY 1136 GGATGGAGTCTGTACCTACAAAACCGTAGTGGCAGAGCTGACCAGACCGTGGGAATCT 11195
|||
Db 113080 GAGGGAGGCTGTACCTGCAAAAGCCACAGGGCAGAGCTGTCAAGACCATGGGAA-CC 113022
|||||

QY 1196 ACCTCTTGCAATGTGATGACCTGGAGCTGGAGCATGGAGTCAAAAGAGATCA---TTT 1252
|||||
Db 113021 ACCTCTTGCAATGAGTAACTGGATGTGAGACATGGAGTCAAAAGGAGATCATTTT 112962
|||||

QY 1253 GAGCTTTAGATTTGACTGCCCCACTGGATTTTCGGACTTATATGGGGCCCGTA-CCCT 1311
|||
Db 112961 GAACTTTAAAGATTTGACTGCCCCACTGGATTTTCAGAGTTGCAGGGGGCCTGTAG 112902
|||||

QY 1312 TGTTTTGGCCAAATTTTTCATTTGGAACCTGCGGTATTTTACCAATGCTGTACCTCCAT 1371
|||||
Db 112901 TGTTTTGGCCAAATTTTCTCGAATTTGGAATGGCTGTATTTTACCAATACCTATAC 112842
|||||

QY 1372 TGTAT 1376
|||||
Db 112841 TGTAT 112837
|||||

RESULT 10
US-09-734-674-3/c
; Sequence 3, Application US/09734674
; Patent No. US20020081648A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001018
; CURRENT APPLICATION NUMBER: US/09/734,674
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 202001
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc\_feature
; LOCATION: (1)...(202001)
; OTHER INFORMATION: n = A,T,C or G
US-09-734-674-3

Query Match 42.9%; Score 590.4; DB 10; Length 202001;
Best Local Similarity 78.4%; Pred. No. 1.e-166;
Matches 803; Conservative 0; Mismatches 171; Indels 50; Gaps 6;

QY 391 GAAGAAATTTCTAAGCAGCAAAAGCATTCAGAGGTGACTTGGGTGCTGTAAAGGCATTC 450
|||||
Db 40027 GAAGAAATTTCTAAGCAGCAAAAGCATTCAGAGGTGACTTGGATGCTGTAAAGCATTC 39968
|||||

QY 451 AGTTTCATAGAGGAGCGACGATAGAGTTTCAGAAAAATTTGCACCCCTGCACAAATGTGATA 510
|||||
Db 39967 GCTTTATAAAGGAAGCAAGGATATAAAGTTTCAGAAAAATTTGCAGCCTGCACATGCAATA 39908
|||||

QY 511 AAAAAAGAAAAACCCATTTTCTGAGGGGAAATTCAGAGTGGCTGCAGAAATTTGCATATGT 570
|||||
Db 39907 GAAAAAGAAATTCATTTTCTGAGGAGAAATTCAGGCCAAATTCAGAAAGT 39848
|||||

QY 571 AATGAGGAGCTGAATGTTAATCCTCAAGACAATGGGAAAAATATCTCCCTGGACATGTCA 630
|||||
Db 39847 AATGAGGAGCAAAATGTTAATCCCAAGACAATGGGAAAAATGCTTCCAGGGCATGTCAA 39788
|||||

QY 631 AGGTCTTTCACAGCAGTCCATCAATCACTGGCTGGAGGCCTAG---GAGAAAAATGGTTT 687
|||||
Db 39787 AGGTCTTATGCGAGCCCTCCCATCACAGACCCAGAGGCCCGAGGAAGAAAAAGTGGTTT 39728
|||||

QY 688 TGTGGGACAGGCCCGGTCCTGTGCTGTGTCAGCCTTAGAGACTTGGTGCCTGTGTC 747
|||

Db 39727 TCTCAGTCAGGCAGAGGTACCCATCCTGTGTACAGCCTAGGAGCTTGGTGCCTGTGTT 39668  
Qy 748 CCAGTTAATTC-----GCTGTGCTTCAGAG 774  
Db 39667 TCAGCTGTCCAGCCATAGCTGAAGGACAAATCTAGAGCTCAGGCTGTGCTTCAGAG 39608  
Qy 775 GGTGAAGCCCAAGCCTTGGAGCTTCAAGTGTGTGAGCCTGTGGTGCAGAGAG 834  
Db 39607 GGTGAAGCCTCAAGCCTTGGAGCTTCCATGTGTGTGAGCCTGCAGGTGTACGGAAG 39548  
Qy 835 TCAAGAAATGAGTTGGGAACCTCCATCAGATTCAGAGATATATGGAACCCCTGG 894  
Db 39547 TCAAGAAATGAGTTGAGAACTCTGCTAGATTTTCAGAGATGATGGAATGCCCTGG 39488  
Qy 895 ATGCCAGGCAAGTTTGTCTAGGGTGGGGTCCCTCATGAGAACCTCTGCAAGGTA 954  
Db 39487 ATGCCAGGCAAAATTTGCTTCAGGGCAGGGCCCTCATGAGAACCTCTGCTAGGCA 39428  
Qy 955 GTACAAAGGGAATGTTGGGTGGAGCCGCCACACAGAGTCCCGAGTGGGCTCCATCT 1014  
Db 39427 GTGCAGAAAGGAATGTGGGTTACAGCCCC--CACAGAGTCCCTACTGGGACACCACT 39370  
Qy 1015 AGTAGAGCTGTGAGAGAGTCCACCATCTCCAGACTCCAGAGGTAGATCCACTGAC 1074  
Db 39369 AGTGAGCTGTGAGAAAGGGCCACTGTCTCTGACTGCAGATGGAAGATCCACTGAC 39310  
Qy 1075 AGCTTGCAGCATGTGCTTGAATAATCCAC-AGACACTCAGTGCCAGCCTGTGAAAGCAGC 1133  
Db 39309 AGCTTGAATGTGTCTCGGAAACACACAGACTCAACGCCAGCCCATGAAAGCAGC 39250  
Qy 1134 AGGATGAGCTGTGACCTTCAAAACCGTAGTGCGAGCTGACCAAGCCGTGGGAAT 1193  
Db 39249 CAGGAGTGGGTATATCTGCAAGCCACGAGGCGAGCTGCCCAAGACCATGAGAA- 39191  
Qy 1194 CTACCTCTTGATCTCATGACCTGGAGCTGAGACATGGATCAAAAGAGATCATTTGG 1253  
Db 39190 -----CCATGACATGACCTGCATGTGAGACATGGATGAAAGAGATCATTTGG 39140  
Qy 1254 AGCTTAAAGATTGACTGCCCACTGGATTTCGGACTTATATGGGGCCCGTA-CCCCTTT 1312  
Db 39139 AGCTTAAAGATTGATGCTGCTGGATTGAGCTTGGACTTGGCCCTGTAGCCCCCTT 39080  
Qy 1313 GTTTGGCCAAATTTTCCATTTGGAATGCGGTATTTACCAATGCTGTACCTCCATT 1372  
Db 39079 GTTTGGCCAAATCTCCCATTTGGAATGGCTGTATTTACCTAATGCTTTACCCAGTT 39020  
Qy 1373 GTAT 1376  
Db 39019 GTAT 39016

RESULT 11  
US-09-263-959-1/c  
; Sequence 1, Application US/09263959  
; Patent No. US20020150891A1  
; GENERAL INFORMATION:  
; APPLICANT: Hood, Leroy E.  
; APPLICANT: Rowen, Lee  
; APPLICANT: Koop, Ben F.  
; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC COMPOSITIONS AND METHODS WHICH UTI  
; NUMBER OF SEQUENCES: 1279  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: US  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/263,959  
; FILING DATE: 05-MAR-1999  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McMasters, David D.  
; REGISTRATION NUMBER: 33,963  
; REFERENCE/DOCKET NUMBER: 920010.426C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 684973 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-09-263-959-1  
  
Query Match 40.2%; Score 553.8; DB 10; Length 684973;  
Best Local Similarity 75.4%; Pred. No. 2.3e-155;  
Matches 797; Conservative 0; Mismatches 217; Indels 43; Gaps 7;  
  
Qy 351 GAATATTTTGGCTAATGAGAGCAAAATTTGGCTAATTAAGGACATTCAGTTTCATTAAGGAGCAGA 410  
Db 560894 GAACTTTTGAACCTTGAGAGAGATGATTAGCGTATCTGGCAGAGAAATTTCTTAAGCAGCA 560835  
Qy 411 AGCATTTCAAGAGGTGACTTGGTCTGCTTAAGGACATTCAGTTTCATTAAGGAGCAGA 470  
Db 560834 AAGTCTTCAAAAGGTGACTCGGGTCTATTAAGACATTTCCATTTTAAAGAGAAACAGA 560775  
Qy 471 GCATAAGAGTTTCAGAAAATTTGCACCTGCACAAATGTGATAAAAAGAAAACCCATTTTC 530  
Db 560774 ACATAAAGATTATTAATTCGTAGCCTGCACANTCCAGTGGAA---AAAAAACCATTTTC 560718  
Qy 531 TGAGGGGAAATTCAGCTGGCTGCAGAAATTTGCATATATGTAATAGGAGCTGAATGTTAA 590  
Db 560717 TGAGGAGAAATTCAGACAGCTGTAGAAATTTGCACA-----AGAAGCCTTAATGTTAA 560665  
Qy 591 TCCTCAACACATGGGAAATATCTCTGGACATGTACAGAGCTTTCACAGCAGTCCAT 650  
Db 560664 TCCCAAAACCATC-GGGAAATGTCTCCAGGCCGTGCAGAGACCTTCATGCAGCCCT 560606  
Qy 651 CAATCACTGGCCTGGAGCCCTAG---GAGAAATGGTTTGTGGGACAGGCCAGGGTC 707  
Db 560605 CCCATTACAGCCTGGAGGTCCAGGATGAAATGTTTATATGGCCAGGCCAGGGTC 560546  
Qy 708 CTTGTGCTGTGTCAGCCTAGAGACTTGTGTCCTGTGTCCAGTTAATTCAGCTGTGGC 767  
Db 560545 CCCATGCCATGTGCAGCCTAGGAGCTTGTGTCTGTGTCCAGCCACTCCAGCCATAGA 560486  
Qy 768 TTCAGAGG-----TGCAAGCCCCAAGCCTTGGCAGCTT 801  
Db 560485 TCAAGGGGCCAACATAGAGCTCAGTTGTGTCTTCAGAGGGTGGAGCCTTGGCACCTT 560426  
Qy 802 CCAAGTGTGTGAGCCTGTGGGTGCAAGAGTCAAGAAATGAGTTTGGGAACTCCA 861  
Db 560425 TTATGTGTGTGAGCCTGTAGGTGCACAGAGTCAAGAAATGAGTTTGGGAACTCCA 560366  
Qy 862 ATCAGATTTCAAGAGATATATGGAACCCCTCGATGCCAGCAGAGCTTTGCTGTAGGG 921  
Db 560365 CCTAGATTTTGAAGATGATGAGAAAGCCCTGAGAGGCTTGGATGCCAGCAAAA-TTTCCTGCAGGG 560307  
Qy 922 GTGGGTCTTCATGAGAACCTCTCAGAGGTAGTAGCAAAAGGGAATTTGGGTGGAG 981  
Db 560306 GTGGGTCTTCATGAGAACCTCTCAGAGGTAGTAGCAAAAGGGAATTTGGGTGGAG 560247  
Qy 982 CCCCACACAGAGTCCCGAGTGGGCTCATCTAGTAGAGCTGTGAGAGAGTCCACCA 1041  
Db 560246 CCCCACACAAATTCCTTACCTGGGTAGCGCTAGTGGAGCTGTGAGAGAGGCGCCACCG 560187  
Qy 1042 TCCTCCAGACTCCAGAGAGGTAGATCCACTGCACACTTGCAGCATGTGCCTGAAATCC 1101  
Db 1101 TCCTCCAGACTCCAGAGAGGTAGATCCACTGCACACTTGCAGCATGTGCCTGAAATCC 1101



Db 14563 AGCAACAGGAACTCAATATTAAATCAACCAAGCAATGGGAAAATGTCTCCAGGCGATGC 14504  
Qy 628 CAGAGGTCTTACAGCAGTCCATCAAAATCACTGCGCTGGAGGCCCTAGAG---AAAATGG 684  
Db 14503 CACAGACCTTTGGCGCAGCCCTACCACTACAGACTCAGAGCCTAGGAGCAAAAATGG 14444  
Qy 685 TTTTGTGGACAGGCCAGGCTCCCTGTGCTGTGTGTGAGCCCTAGAGACTTGGTGCCCTGT 744  
Db 14443 TTTATGGGCGAGGTCACAGGTCCTCTGCTGTGTGCTGCTGAGGACTTCTGCTACCTGC 14384  
Qy 745 GTCCAGTTAATTAC-----CTGTGCTTCA 771  
Db 14383 ATCCAGCTGCTCAGCCATGACTGAAGGGGCCAAGTACAGCTCAGGCTGTGCTTCA 14324  
Qy 772 GAGGTGCAAGCCCCAGCCTTGGCAGCTTCCAAGTGGTGTGAGCCCTGGGGTGCAAG 831  
Db 14323 GAGGTACAAGCCCCAAGCTGTGCAGCTTCCATGTGTGTGAGCCGCGAGGTGCACAG 14264  
Qy 832 AAGTCAAGAAATGAGTTTGGGAACCTCCAATCAGATTTTCAGAGATATATATGGAACCCC 891  
Db 14263 AAGTCAAGAAATGAGTTTGGGAGCCTCTGCTGCAATTTTCAGAGGATGTATGGAACACC 14204  
Qy 892 TGGATGCCAGCGCAGAACTTGTCTAGGGTGGGTCTCATGAGAACCTCTGCAAGG 951  
Db 14203 TGGATGCCAGCGAAGTTTGTGAGGGTGGGGCTCATGAGAACCTCTGCTAGG 14144  
Qy 952 GTAGTACAAAAGGAAATGTTGGGTGGGAGCCCCACACAGAGTCCCGAGTGGGGCTCCA 1011  
Db 14143 GCAATGCAAGAGGAAATGTGGGTGGTGGCCCTATACAAAGTCCCGACTGGGCACTG 14084  
Qy 1012 TCTAGTAGAGCTGTGAGAAGTCCACCATCTCCAGACTCCAGAGGGTATGATCCACT 1071  
Db 14083 CCTAGTGGAGCTCTGAGAAGAGGGCCACCATCTCTACACCCAGCAATGATATATCCACC 14024  
Qy 1072 GACAGCTTGACAGTGCCTTGAATAATCCACAGACACTCAGTGCAGCCCTGTGAAAGCA 1131  
Db 14023 AACAGCTTGACCATGCACTTGGAAAGCCACAGACACTCAATGCCAGCCCCACGAAAGCA 13964  
Qy 1132 GCAGGGATGGATCTGTACCTACCAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGA 1191  
Db 13963 GCCGGGAAAGGCTGTACTCTGCAGAGAGCCACAGGGCAGAGCTGCCTAAGACCATGGGA 13904  
Qy 1192 ATCTACCTCTTGCATTTGCATGACCTGGACGTGAGACATGAGATCAAAAGATCATTTT 1251  
Db 13903 ACCCAACTCTTGCATCAGCATGACCTGGATGTGAGATAGATCAAAAGATGTTCAATTT 13844  
Qy 1252 GGAGCTTTAAGATTTGACTGCCACCTGATTTGCGACTTATATGGGGCCGTA-CCCT 1310  
Db 13843 GAAGCTTTAAGATTTGACAGCCCTGCTGGATTTTCAGACTTGCATGGGGCCCTGTAGCCCT 13784  
Qy 1311 TTGTTTGGCCAAATTTTTCATTTGGAACCTGCGGTATTTACCAATGCCCTGTACCTCCA 1370  
Db 13783 TTGTTTGGCCAAATTTCTCCATTTGGAATGGCTGAATTTACCAATGCCCTGTACCCCTA 13724  
Qy 1371 TTGTAT 1376  
Db 13723 TTGTAT 13718

RESULT 14  
US-09-864-761-2925  
; Sequence 2925, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Acomlca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 2925  
; LENGTH: 1946  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC006337.2  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 14  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 15  
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 13  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 12  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 12  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 12  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 10  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 14  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 13  
US-09-864-761-2925

Query Match 38.0%; Score 523; DB 10; Length 1946;  
Best Local Similarity 74.4%; Pred. No. 1.5e-147;  
Matches 764; Conservative 0; Mismatches 180; Indels 83; Gaps 5;

Qy 351 GAATATTGGCTAATAGAGAGCAAAATTTGGCTAATAGGAGAAATTTCTTAAGCAGCA 410  
Db 174 GAATTTTGAACCTGAAAGAGATGATTAGGGTATCTGGTGAAGAAATTTCTTAAGCAGCA 233  
Qy 411 AAGCATTCAGAGGTGACTTGGTCTCTTAAGGCATTCAGTTTCATAAGGGAGGCAGA 470  
Db 234 AAGCATTCAGAGGTGACTTGGTCTCTTAAGGCATTCAGTTTCATAAGGGAGGCAGA 293  
Qy 471 GCATAAGAGTTTCAGAAAATTTGCACCCCTGACAACTGTATAAAAAAGAAAAACCCCATTTTC 530  
Db 294 GCATAAAGTTTCAGAAAATTTGCACCCCTGACAACTGTATAAAAAAGAAAAACCCCATTTTC 353  
Qy 531 TCAGGGAAATTCAGAGCTGGCTGCAGAAATTTGCATATGTATATGAGAGGAGCTGAATGTTAA 590

```
Db 354 TGGAGGAAATTCAGAGTCCCTGCGAGAAATTTGCAATAAGTACGAGGAGCCCAATGTTAA 413
QY 591 TCTCTAAGACAATGGGAAAATATCTCTGGACATGTCTAGAGTCTTACAGAGTCCAT 650
Db 414 TCCCAAGACATGAGAGAAATGTCTCCAGGCATGTCTAGAGTCTTCA-GCAGCCCTT 472
QY 651 CAAATCACTGGCTGGAGCCCTAGGAGAAATGTTTGTGGAGAGGCCGCCAGGTCCCT 710
Db 473 CCATCGCAGGCCTGGATGCCCTAGGAGAAATGGTTTTTATAGGCCTGGCCAGGGTCCC 532
QY 711 GTGCTGTGTGACGCTAGAGACTTGGTCCCTGTCTCCAGTTAATTCAGCTGTGCTTC 770
Db 533 GTGCTGTGTGACGCTAGGATTTGGTGTACTGCACTCTACCACTCCAGCCATGGCTGA 592
QY 771 AGAGGTGCAAGCCCAAGCCCTTGCAGCTTCCAAATGGTGTGAGCTGTGGGTGCAAA 830
Db 593 AAGGG-----ACAAATGTAGAGCTTGAAGCCTGTGAGTCCAG 628
QY 831 GAAGTCAAGAAATGAGTTTGGGAACCTCCAAATCAGATTTTCAGAAATATATGAAACCC 890
Db 629 GGGTGCAAGCCCAAGCTTTGGCAGCTTCCA----- 659
QY 891 CTGGATGCCCGCAGAGATTTGCTGTAGGGTGGGTGCTCATGTGAGAACTCTGCAAG 950
Db 660 -----CGTGGTGGGACGCTCATGTGAGAACTCTGCTAG 692
QY 951 GGTAGTACAAAAGGAAATGTTGGTGGGAGCCCAACACAGAGTCCCAAGTGGGCTCC 1010
Db 693 GCGAGTGCAGAGGAATGTGGGTGGAGGCTCCACAGAGTTCCTACTGGGGCACT 752
QY 1011 ATCTAGTAGAGCTGTGAGAAGAGTCCACCATCTCTCCAGACTCCAGAGGTAGATCCAC 1070
Db 753 GCGTAGTGGAGCTTTGAGAAGAGGCCACCATCTCCAGACCCAGATGATAGTCAA 812
QY 1071 TGACAGCTTGACGATGTGCTGAAAATCCACAGACTAGTCCAGCCAGCTGTGAAAAC 1130
Db 813 CAATAGCTGTGACTACGTGCGCTGSAAGAACCCACAGACACTCAACCGCAGCTGTGAAAAC 872
QY 1131 AGCAGGATGGAGTGTGTAACCTCAAAACCGTAGTGCAGAGCTGACCAAGACCGTGG 1190
Db 873 AGCTGGGAGGTGGCTGTACCTGCAAGCCATAGGGTGGAGCTGCCAAGACCATGG 932
QY 1191 AATCTACCTTTCATGTGATGACCTGGAGCTGAGACATGGAGTCAAAAGAGATCATTT 1250
Db 933 AACCACCTCTTGCAATTTGCATGACCTGGATGTGAGACCTGGAGTCAAAAGAGATAATTT 992
QY 1251 TGGAGCTTTAAGATTTGATGCCCCACCTGGATTTCCGGACTTATATGGGGCCGT-ACCCC 1309
Db 993 TGGAGCATTAAGATTT-ACTGCCCCCTCTGGATTTTCAGACTTGCATAGGCCCTGTAAACCCC 1051
QY 1310 TTTGTTTGGCCAAATTTTTCATTTGGAACCTGCGGTATTTACCAATGCGCTGTACCTCC 1369
Db 1052 TTTGTTTGGCCAAATTTTCCCATTTGGAATGGCTATATTTATCCATACCTGTATCCCC 1111
QY 1370 ATTGTAT 1376
Db 1112 ATTGTAT 1118
```

RESULT 15  
US-10-254-869-3  
; Sequence 3, Application US/10254869  
; Publication No. US20030027307A1  
; GENERAL INFORMATION:  
; APPLICANT: YE, Jane et al  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; FILE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CL001160DIV  
; CURRENT APPLICATION NUMBER: US/10/254,869  
; CURRENT FILING DATE: 2002-09-26  
; NUMBER OF SEQ ID NOS: 8

```
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 148567  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc_feature  
; LOCATION: (1)...(148567)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-254-869-3
```

```
Query Match 36.5%; Score 502.8; DB 9; Length 148567;  
Best Local Similarity 71.5%; Pred. No. 2.4e-140;  
Matches 781; Conservative 0; Mismatches 232; Indels 79; Gaps 6;  
  
QY 352 AATATTGGCTAATGNGGNAAGCAAAATTTTGGCTAATAAGGAAGAATAATTTCTAAGCAGCAA 411  
Db 100932 AACTTTGAACTTGAGAGAGATAATTTAGGGTATCTGTATGGAAGAATAATTTCTAAGCAGCAA 100991  
  
QY 412 AGCATTCAAGAGGTGACTTGGGTGCTTTAAAGGCATTCAGTTTCAT----- 458  
Db 100992 AGCAITCAAGAGGTGACGTGGGTGCTTTAAAAACATTAAGTTTATTTCATTTCACAAAGA 101051  
  
QY 459 -----AAGGAGGCGCAGCATTAAGAGTTTCAGAAAAAT 489  
Db 101052 TATGTTTGGAAATTAGAACTCATGTTTAAAGAAAAGCAGGAATAAAAGTTTCAGAAAAAT 101111  
  
QY 490 TTGACACCTGCAAAATGTATATAAAGAAAACCCATTTTCTGAGGGAATAATTCAGCTG 549  
Db 101112 TTATAGCTGATGATGGAATAAGAAAAGAAAACCTTATTTCTGAGGAGAAAATTTCAAACATG 101171  
  
QY 550 GCTGCAGAAAATTTGCATATGTAAATGAGGAGCTGAATGTAAATCCCTCAAGACAATGGGAA 609  
Db 101172 GCTGCGAAATTTGCATCAGTAATCAGGAGCAAAATGTTAATGGCCAGACCATGGGAA 101231  
  
QY 610 AATATCTCTTGACATGTCTAGAGTCTTTCACAGAGTCCATCAATCACTGCGCTGGAGG 669  
Db 101232 AATGTCTCCAGGCGATGTCAGAGGT-----AGCCCCCTCTATCAACAAGCCCTGAGT 101282  
  
QY 670 CCTAGGA-GAAATGTTTGTGGCAGCAGGCCAGGCTCCCTGTGCTGTGTGACGCTAG 728  
Db 101283 CTTGGGAGAAAATATGTTTATGCGGTGGGCCCGGCGCTTGTGCTTTCGTAGTCTCA 101342  
  
QY 729 AGACTTGGTGGCTGTGCCAGTTTAAATTCAGCTGTG----- 765  
Db 101343 GGACTTGTGCTGCTCATCCAGCTGTTTCTAAAGGGGCCAATACAGTTTCAGACCATTT 101402  
  
QY 766 GCTTCAGAGGTTGAAGCCCAAGCCCTTGGCAGCTTCCAAAGTGTTCAGAGCTGTGGGT 825  
Db 101403 GCTTCAGAGGTTGAAGCAGCAAGCCCTTGGTGGCTTACGCATGTTGTGGGCTGTGGAT 101462  
  
QY 826 GCAAGAAAGTCAAGAAATTTGAGGTTTGGGAACCTCCAAATCAGATTTTCAGAAGATATATGA 885  
Db 101463 GCACAGAAAGTCAAGAAATTTGAGGTTTGGGAACCTCTGCGCTGGATTTTCAGAGGATGATGA 101522  
  
QY 886 AACCCCTGGATGCCAGCAGCAAGTTTGTCTGATAGGGTGGGTCTCATGGAACCTCT 945  
Db 101523 AATGCTAGATGTCCGACAGAGTTGTCTACATGGCAGAGGCCCTTATGGAGAACCTCT 101582  
  
QY 946 GCAAGGTTAGTACAAAAGGAAATTTGGTGGGAGGCCCAACAGAGTTCGCCAGTGGG 1005  
Db 101583 GCTAGGGCAGCGTGAAGGAAATATGGGTGGGAGAACCCACACAGAGTTTCCCAGTAG 101642  
  
QY 1006 GCTCCATCTAGTAGAGCTGTGAGAAGAGTCCACCATCTCCAGACTCCAGAAGGTTAGA 1065  
Db 101643 GCACCACCTAGTGGAGCTGTGAGAAGAGTCCACCATCTCCAGACACCAAGTGTAGC 101702  
  
QY 1066 TCCACTGACAGCTTCAGCATGTGCTGAAAATCCAGACACACTCAGTCCAGGCTGTG 1125  
Db 101703 TCCACCAACAGTTTGCACCATGTGCTGAAAAGCTGCAGACATACATGCCACCAATG 101762  
  
QY 1126 AAGCAGGAGGATGGATGCTGTACCCCTACAAAACCGTAGTGGCAGAGCTGACCAAGACC 1185  
Db 1126 AAGCAGGAGGATGGATGCTGTACCCCTACAAAACCGTAGTGGCAGAGCTGACCAAGACC 1185
```



THIS PAGE BLANK (USPTO)



Db 47617 GGAATAATGGTTTGTGCTGGGGCCCGAGGCGCTTGTGTTGTGCGAGTCTCAGGACT 47676  
QY 734 TGGTCCCTGTGTCAGTAATTAATTCAGCTGTGGCTTCAGAGGGTGCAAGCCCAAGCCCTT 793  
Db 47677 TGGTCCCGACATCCAG-----CAGTGGCTA 47703  
QY 794 GGCAGCTTCCAAAGTGGTGTGAGCCTGTGGTGGTGAAGAAGTCAAGAAATGTAGGTTTGG 853  
Db 47704 AAAGGGGCAATGTACAGCTTAGACCTTTGCTTCAGAGGTG-CAAGCCCAAGCCCTGGT 47762  
QY 854 AACCTCCATCAGATTTTCAGAGATATATGGAACCCCTGGATGCCAGGCGAAGATTGG 913  
Db 47763 GGCCTACATGTGTTGGGCTGCAGAT-----ACACAGAAATTTG 47804  
QY 914 CTGTAGGGTGGGCTCTCATGGAAACCTCTGCAAGGTAGTACAAAGGGAATGTTG 973  
Db 47805 CTGCAGTGGTGAACCTCATGTAGAACCTCTGCTAGGCGAGTGTAGAAGTATATGG 47864  
QY 974 GGTGGAGCCCGCACAGAGTCCCAAGTGGG-----CTCCATCTAGTAGAGCTGTGA 1027  
Db 47865 GGTGGAGCCCGCCACACAATCCCACTGGGGCACTGCCTACTGCTACTGGAACCTGA 47924  
QY 1028 GAAGAAGTCCACCATCTCCAGACTCCAGAAGGTAGATCCACTGACAGCTTGCAGCATG 1087  
Db 47925 GAAGAAGGCCACCATCTCCAGACCCCAAGATGGTAGATCCACTGATGGCTTGAACCATG 47984  
QY 1088 TGCTGAAAAATCCACAGACACTAGTGCAGCCTGTGAAAGCAGCAGGGATGGAGTCTG 1147  
Db 47985 CACCTGGAAAGCCACAGACACTCAACACCGCTGTGAGGCGACTGGAAGGAGGCTG 48044  
QY 1148 TACCCTCAAAAACCTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTTTTGCATT 1207  
Db 48045 TACCCTGCAAAAACAGAGGAGAGCTGCCCAAGGTGATGGGACCCACCTCTTGCATG 48104  
QY 1208 GTCATGACCTGGACGTGAGACATGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTG 1267  
Db 48105 AGCCTGACTTGAATGTAGACATGGAGTCAAGAGGAGATCATTTTGGAGCTTTAAGTTTG 48164  
QY 1268 ACTGCCCCACTGGATTTGGAGCTTATATGGGCCCGT-ACCCCTTTGTTTGGGCCAATTT 1326  
Db 48165 ACTGCCACCTGGATTTGGAGCTTGCATGGGCCCTGTGGCCCTTCATTTTGGCCAAATTT 48224  
QY 1327 TTTCCATTTGGAACCTGCGGTATTTACCAATGCCGTGACCTCAATGTAT 1376  
Db 48225 ATCCCATTTGGAATGGGTATATTTACCAATGCCGTGACCCCAATCTAT 48274  
RESULT 2  
US-09-146-053-5  
; Sequence 5, Application US/09146053A  
; Patent No. 6399349  
; GENERAL INFORMATION:  
; APPLICANT: Ryan, James W.  
; APPLICANT: Sprinkle, Terry Joe Curtis  
; APPLICANT: Venema, Richard C.  
; TITLE OF INVENTION: Human Aminopeptidase P Gene  
; FILE REFERENCE: MC0103  
; CURRENT APPLICATION NUMBER: US/09/146,053A  
; CURRENT FILING DATE: 1998-09-02  
; EARLIER APPLICATION NUMBER: 60/057,854  
; EARLIER FILING DATE: 1997-09-02  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 44453  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-146-053-5

Query Match 20.5%; Score 282.8; DB 4; Length 44453;  
Best Local Similarity 62.2%; Pred. No. 1.8e-79;  
Matches 585; Conservative 0; Mismatches 332; Indels 23; Gaps 8;

QY 444 GGCATTTCAGTTTCATAAGGGAGGCGAGCAGATAGAGTTTCAGAAAAATTTGCACCTGACAA 503  
Db 38135 GGAATTTTATTTAATGGAGCAGAGTGTAAAATTTGGAAAAATTTGCACCTCACA 38194  
QY 504 TGTGATAAAAAGAAAAACCCATTTTCTGAGGGAAAAATTTCAAGCTGGCT-----GCAGAA 558  
Db 38195 TGTGATAGAAAAAGAAAAAGCCAAATTTTCAGGGGATGAATTTCAAGCAGACTGTGGGTAAAC 38254  
QY 559 ATTTCATATGTAAATCAGAGCTGAATGTAAATCTCCTCAAGACAAATGGGAAAAATATCTCC 618  
Db 38255 ACTTGTAGAGATATTTGCATAACTAAAAAGACATTTTCAGAGATCTAAGAGCAGCCCC 38314  
QY 619 TGGACATGTTCAGAGGTCTTTCACAGCAGTCCATCAATCACTGCCCTGGAGGCTTAGAGA 678  
Db 38315 TCCCATCAGAGCTCTGAGGCCAAGTGTTCATGGTCCAGGCTTAGGCCCTGTCTGCC 38374  
QY 679 AAATGGTTTGTGGGACAGGCCAGGGTCCCTGTGTGTGTCAGCAGCTAGAGACTTTGGTG 738  
Db 38375 CTGAGCAGCCTCTGGACACTGCTTCTCGCATCTCTGCCCACTCCAGCTCCAGCCTTGCTG 38434  
QY 739 CCTGTGTCCAGTTAA--TTCAGCTGTGCTTTCAGAGGTCGAGGTCGCAAGCCCTTGGC 796  
Db 38435 AAAGTCCCCAGATACAGCTCAGTTTGTCTTCAGAAAAGTGTAGCTGTACGCCCTTGGC 38494  
QY 797 AGCTTCCAAAGTGGTGTGAGCCTGTGGGTGCAAGAAAGTCAAGA---ATTGAGTTTGGG 853  
Db 38495 AGCTTCTGTCTGCATTAAGCCTGTAGGTGTCAAAAGTCAAGAGTGAAGAGGCTTGGC 38554  
QY 854 AACCTCCAATCAGATTTTCAGAGATATATGGAACCCCTGGATGCCAGGCGAAGATTGG 913  
Db 38555 AGCCTCTGCTAGATTTCAGAAAGATGTATAGGAA-ATCTGGGTGCCCAGGCAAG-AGCCTG 38612  
QY 914 CTGTAGGGTGGGTCTCATGAGAACCTCTGCAAGGCTAGTACAAAGGGAATTTG 973  
Db 38613 CTACAGGGGCGAGGCCCTCATGGAAAACCTCTGTAGGACAGTGCAGAGGGGAAAGCTGG 38672  
QY 974 GGTGGAGCCCCACACAGAGTCCCCAGTG-GGGTCCATCTAGTAGAGCTGTGAGAAGA 1032  
Db 38673 GGTGGAGCCCCCATGTACAGTCTCCACTGAGGCACTCCCTAGTGGGCTGTGGGAAGG 38732  
QY 1033 ACTCACCATCTCCAGACTCCAGAGGTAGATACACTGCAGCTTGCAGAGTTCGAGCATGTCCT 1092  
Db 38733 GGGCCCCGGTCTCCAAACCTCAGAATGTAGATCCACCAAGCTCGCACCTGCACT 38792  
QY 1093 GAAATATCCAGACACTCAGTCCAGCCTGTGAAGCAGCAGGGATGGAGTCTGTACCC 1152  
Db 38793 GGAAGAGCCACAGGCACTCAACAATG-----TGAAAGCAGCTACAGGTGCCGAACCC 38845  
QY 1153 TACAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTCTTGCATTGTCTAT 1212  
Db 38846 AGAAAAGCTACAGGGCGCAGAGTGCCTGAGGCTTGGGAGCCCAACCTTGTGTCACTGT 38905  
QY 1213 GACCTGGAGCTGAGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTTAAGATTT---GAC 1269  
Db 38906 GGCCTGGATGTGAGACCTGGAGTTGAAGGAGATTAATTTGGAGCTTTTGACATTTAAAGAC 38965  
QY 1270 TGCCTCAGTGGATTTCGGACTTATATGGGCCCGTACCCCTTGTGTGGCAAAATTTT 1329  
Db 38966 TACCCTGCTGGGTTTCAAACTTGCATGGGCCCTGTAGCCCTTATTTTGGCCAAATTTTC 39025  
QY 1330 CCATTTGGAATGCGCTATTTTACCCCAATGCCGTGACCTCC 1369  
Db 39026 CCTTTTGAATGAAATGTTTACTCAATGCCCTATCCCCC 39065  
RESULT 3  
US-08-687-080-59/c  
; Sequence 59, Application US/08687080  
; Patent No. 5965427  
; GENERAL INFORMATION:  
; APPLICANT: Gregory Dolganov  
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof  
; NUMBER OF SEQUENCES: 175

```

CORRESPONDENCE ADDRESS:
; ADDRESSEE: Denlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,080
; FILING DATE: 17-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/592,126
; FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111.30
TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
LENGTH: 14855 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 5' END OF INTRON 2 OF RAD50 GENOMIC
; INDIVIDUAL ISOLATE: SEQUENCE
US-08-687-080-59

Query Match          19.7%; Score 271.2; DB 2; Length 14855;
Best Local Similarity 77.3%; Pred. No. 4.9e-76; Indels 15; Gaps 4;
Matches 385; Conservative 0; Mismatches 98;

QY   890 CCTGGATGCCAGCAGAAGTTTCTGTAGGGTGCGGTCCTCATGGAGAACCTCTGC AA 949
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   14854 CTGTGATGCCAGGCACATGTTGCTCAGGGTGCGGCGCCCTTATGGA AACCTCTGCTA 14795

QY   950 GGGTAGTACAAAAGGGAATGTTGGGTGGGAGCCCCCACACAGAGTCCCAGTGGGGCTC 1009
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   14794 GGGCAATATGGAAGGAAATGTGGGTGTAACCCC--ACAGAGTTCCTATGGAGGGGAC 14737

QY   1010 CATCTAGTAGAGCTGTGAGRAGAGTCCACCATCTCCAGACT-----CCAGAAG 1059
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   14736 TGCCTAGTGGAGCTGTGAGAAGACAGCCACTGTCTCCAGACTGGTATAGATCCCCCAG AAT 14677

QY   1060 GGTAGATCCACTGACAGCTTGCAGCATGTGCTGAAAAATCCACAGACACTCAGTGC GCAG 1119
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   14676 ATAGATCCACTGACACTTGCACATGTGCACCTGGAAAACTGGAGGCACATCAAC ACCAG 14617

QY   1120 CCGTGTGAACAGCAGGGAGTGGAGTCTGTACCCCTACAAAACCCGTAGTGGCAGAGCTG ACC 1179
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   14616 CCTGTGAACAGCCAGGAGGAGGAGGCTATACCCCTGCAAAGCC--AGAAGTGGAGCTGCC CC 14559

QY   1180 AAGACCGTGGGAATCTACCTCTTGCATGTGCATGACCTGACAGCTGAGACATGGAGTCA AA 1239
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   14558 AAGGCCATGGAAGCCCACTCTTGCATCAGAGTGCACCTGGATGTGAGACATGGAGTCA AA 14499

QY   1240 AGAGATCATTTTGGAGCTTTAAGATTTTGAAGTTTGAAGTTTGAAGTTTGAAGTTTGA AGTT 1299
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   14498 GGAGCATCATCTGGAGCTTTAAGATAACACCTTGCCCCCACTGAAATTTTCGGAGTTGC ACAGGG 14439

QY   1300 CCCGTA-CCGCTTTGTTTTGGCCAATTTTTTTCATTTTGGAACTGCCCTGATTTACCCCA ATG 1358

```

[illegible]

RESULT 5  
US-09-062-451-29  
; Sequence 29, Application us/09062451  
; Patent No. 6344550  
; GENERAL INFORMATION:  
; APPLICANT: Fridakis, Tony N.  
; APPLICANT: Smith, John M.  
; APPLICANT: Reed, Steven G.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF BREAST  
; CANCER

Query Match 18.0%; Score 247.4; DB 4; Length 620;  
Best Local Similarity 68.6%; Pred. No. 3e-69;  
Matches 424; Conservative 0; Mismatches 177; Indels 17; Gaps 6;

Qy	771	AGAGGGTGCAAGCCCAAGCCTTGGCAGCTTCCAAGTGGTGTTCAGCGCTGGGGTGCACA	830
Db	2	AGAGGGTACCAGCCCAAGCCTTGACAACCTTCCATAGGGTGTCAAGCCTTGTGGGTGCACA	61
Qy	831	GAAGTCAAGAAATTTGAGGTTTGGGAACCTTCCAATCAGATTTTCAGAAATATATGCA	885
Db	62	GAAGTCAAAAATTTGAGTTTGGGATCCTCAGCCTAGATTTTCAGAGGATATAAAGAAACAC	121
Qy	886	- -AACCCCTGATGCCAGCAGAGATTTTGCCTAGGAGTGGGGTCCCTCATGAGAAACCT	943
Db	122	CTAACACCTAGATATTTCAGACAAAAGTTTACTTACAGGATGAAGCTTTTCACGAAAACCT	181
Qy	944	CTGCAAGGGTAGTACAAAAGGAAATGTTGGGTGGGAGCCCCCACAGAGTCCCCAGTG	1003
Db	182	CTACTAGGAAAGTACAGAGAGAAATGTGGGTTTGGAGCCCCCAAACAGATCCCTCTA	241
Qy	1004	GGGTCCATCTAGTAGAGCTGTGAGAAGATGCCACCATCTCCAGACTCCAGAGGGTA	1063
Db	242	GAACACTGCCCTAATGAAACTGTGAGAAGATGGCCACTGTCTATCCAGACACCCAGAATGATA	301
Qy	1064	GATCCACTGACACTTTCGACGAT -GTGCCTGAAAATCCACAGACACTCAGTGCCAGCC -	1121
Db	302	GACCACCAAAAACCTTATGCCATATGGCTATATAAACCTCAGACACTCAATGCCAGCCC	361
Qy	1122	--TGTGAAGCAGCAGGATGGAGTCTGTACCCCTACAAAACCGTAGTGGCAGAGCTGACC	1179
Db	362	CATGAAAAAAAACCTCAGAGAGAGACTGTNCCCTACATGCCACCGGAGAGACTGCC	421
Qy	1180	AAGACCGTGGGAATTCACCTCTTTCGATTTGTCATGACCTGGAGC -TGAGACATGGAGTCAA	1238
Db	422	CAGGCCATGGAAGCACAGCTCTTATATCAATGTGACCTGGATGTTGAGACATGGAATCCN	481
Qy	1239	AAGAGATCATTTTGGAGCTTTAAGATT ---TGACTGCCCCACTGGATTTCCGACTTATA	1294
Db	482	ANGAAATFCNTTTTAAANACTTCCACGGTTNNAATGACTGCCCTATTANATTCNGAACTTANA	541
Qy	1295	T-GGGGCCGTACCCCTTGTGTTTGGCCAAATTTTTCATTTGGAACCTGCGGTATTTACC	1353
Db	542	TCCNGGCGTGTGACCTCTTGTCTTTGGCCATTTCCCCCTTTTGAATGGCTNTTTTTTC	601
Qy	1354	CAATGCGCTGTACCTCCAT 1371	
Db	602	CCATGCGCTGTNCCCTCTT 619	
RESULT 6			
US-09-598-326-29			
; Sequence 29, Application US/09598326			
; Patent No. 6423496			
; GENERAL INFORMATION:			
; APPLICANT: Frudakis, Tony N.			
; Reed, Steven G.			
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE			
; TREATMENT AND DIAGNOSIS OF BREAST CANCER			
; NUMBER OF SEQUENCES: 247			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Seed Intellectual Property Law Group PLLC			
; STREET: 701 Fifth Avenue, Suite 6300			
; CITY: Seattle			
; STATE: Washington			
; COUNTRY: USA			
; ZIP: 98104-7092			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/09/598,326			
; FILING DATE: 20-Jun-2000			
; CLASSIFICATION: <Unknown>			
; ATTORNEY/AGENT INFORMATION:			

```
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 210121.419D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 620 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 29:
US-09-598-326-29

Query Match      18.0%; Score 247.4; DB 4; Length 620;
Best Local Similarity 68.6%; Pred. No. 3e-69;
Matches 424; Conservative 0; Mismatches 177; Indels 17; Gaps 6;

Qy 771 AGAGGGTGCAGGCCCAAGCCCTGGCAGCTTCCAAGTGGTGTTCAGGCTGTGGGTGCAAA 830
Db 2 AGAGGGTACCAGCCCAAGCCCTGCACAACTTCCATAGGGTGTCAAGCTGTGGGTGCACA 61
Qy 831 GAAGTCAAGAAATTCAGGTTTGGGAACCTCCCAATCAGATTTTCAGAAATATATGA---- 885
Db 62 GAAGTCAAAATTCAGTGTGGGATCCTCAGCCCTAGATTTTCAGAGGATATATAAGAAACAC 121
Qy 886 --AACCCCTGGATGCCAGGCAGAGTGTGCTGTAGGGTGGGTCTCTCATGGAGAACCT 943
Db 122 CTACACCTAGATATTCAGACAAAGTTTACTACAGGGATGAAGCTTTCACGGAAACCT 181
Qy 944 CTGCAAGGGTAGTACAAAGGAAATGTGGGTGGGAGCCCAACACAGAGTCCCCAGTG 1003
Db 182 CTACTAGAAAGTACAGAAAGAAATGTGGGTTGGAGCCCAACACAGAAATCCCTCTA 241
Qy 1004 GGGCTCCATCTAGTAGAGCTGTGAGAAGTCCACCATCTCTCAGACTCCAGAGGGTA 1063
Db 242 GAACACTGCCTAATGAAGCTGTGAGAAGTGGCCACTGTCTATCAGACACACCAAGATGATA 301
Qy 1064 GATCACTGACAGCTTGCAGACT--GTGCTGAAAAATCCACAGACACTCAGTGCCAGCC- 1121
Db 302 GACCCACCAAAACTTATGCCATATTGCCTATAAACTCAGACACTCAATGCCAGCCC 361
Qy 1122 --TGTAAAGCAGCAGGATGGAGTGTGTACCCCTACAAACCCGTAGTGGCAGACTGACC 1179
Db 362 CATGAAAAAAACTGAGAAGAGACTGTNCCCTACAAATGCCACCGAGCAGAACTGCC 421
Qy 1180 AAGACCTGGGAATCTACCTTTCATGTCTACCTGACCTGGAGC--TGAGACATGGAGTCAA 1238
Db 422 CAGGCCATGGAAGCACAGCTCTTATATCAATGTGACCTGGATGTGAGACATGGAATCCN 481
Qy 1239 AAGAGATCATTTTGGAGCTTTAAGATT---TGACTGCCCCACTGGATTTTCGGACTTATA 1294
Db 482 ANGAATCNTTTTAAACTTCCACGGTTNATGACTGCCCTATTANATTTCGAACTTANA 541
Qy 1295 T-GGGGCCCTGACCCCTTGTGTTTGGCCAAATTTTCCATTGGAAGTCCGCTATTTACC 1353
Db 542 TCCNGGCTGTGACCTCTTGTGTTTGGCAATTCCTCCCTTTTGGAAATGGCTTTTTTTC 601
Qy 1354 CAATGCTGTACCTTCCAT 1371
Db 602 CCATGCTGTNCCCTCTT 619

RESULT 7
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
```

```
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match      17.4%; Score 240.2; DB 3; Length 152331;
Best Local Similarity 59.4%; Pred. No. 1.7e-65;
Matches 572; Conservative 0; Mismatches 320; Indels 71; Gaps 7;

Qy 444 GGCATTTCAGTTTCATAGGGAGGCAGACGATTAAGAGTTTCAGAAAAATTTGCACCCCTGACAA 503
Db 71252 GGAACCTATGTTTAAAAATGGAAGTAGAGTCTAAAAATTTGGAAAAATTTGCAGCCCTGGCCT 71311
Qy 504 TGTGATAAAAAAGAAAAACCCATTTTCTGAGGGGAAATTCAGAGCTGGCTGCAGAAAAATTTG 563
Db 71312 TGTGGCAGAGAAAGAAATCCAAGTAGGCTGCAGAGCAATCATTTGCTAG----AGAGATTAG 71367
Qy 564 CATATGTAATGAGGAGCTGAATGTTAATCTCTCAAGACATGGGAAAAATATCTCTCTGCAC 623
Db 71368 CATGACTAAAAAGGGAGGCCAAGTGTCTAATATTCAGACAAATGTTAAAAAGGCTTGGAGGC 71427
Qy 624 ATGTGACAGAGTCTTCACAGCAGTCCATCAATCACTGAGCTGGAGCCCTAG-----GA 676
Db 71428 ATTTTCAGAGATCTATGAAGCAGCCCCCTCCCATCAGAGTGCAGAGGTTTGGTGCACATAGG 71487
Qy 677 GAAATGGTTTTTGGGACAGCGCCAGGGTCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 736
Db 71488 CCCAGAGGTTTTATGGCCANNCCAGGCCACACTGCTATGCACAGCTTTGGGACACTG 71547
Qy 737 TGCCCTGTGTCCCAAGTTAA----- 755
Db 71548 CTGCCCGCATCCAGGCCACTCTGTCTGGCTCCACCCCTTGGCTCAAAGCGGCCAAGATAG 71607
Qy 756 ---TTCAGCTGTGCTTCAGAGGTTGCAAGCCCAAGCCCTTGGCAGCTTCCAAGTGGTGT 812
Db 71608 AGCTTGGACCACTGCTCCCGAGGGCACAAGCCATAAGCCCTTGGTGGTTTCCATGTGGTGT 71667
Qy 813 TGAGCTGTGGGTGCAAGAAAGTCAAGAAAT--TGAGGTTTGGGAACCTTCCAATCAGATTT 870
Db 71668 TAAGCTGCAGTGCACAGATGCAAGATTGAGGGAGCTTGGGCACTTCCACCTAAATTT 71727
Qy 871 CAGAAGATATATGAAAAACCCCTGGATGCCAGGCAAGAAATTTGTGTATAGGGGTGGGGTCC 930
Db 71728 CAGAGGATGTCTCAAAAAACCCCTAGGTTCCAGGCAAGAAAGCATGATACAGGGGCGAGAGCC 71787
Qy 931 TCATGGAGAACTCTGCAAGGGTAGTACAAAAGGAAATTTGGTGGGAGAGCCGCCACAC 990
Db 71788 TTGCAGAGAACTCTACTAGGGCAATGCCAAGGAAATTTGGGGTTGGAGTCTCTCACAC 71847
Qy 991 AGAGTCCCGAGTGGGCTCCATCTAGTAGAGTGTGAGAAGAAAGTCCACCATCTCTCCAGA 1050
Db 71848 ATGGTCCCGACTGGGCACTACCTGGTGATCTGTGGGAATGGGGCTGCTGCCCTCCAGA 71907
Qy 1051 CTCAGAGAGGTAGATCCACTGCAGCTTGCAGCATGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1110
Db 71908 CCCAGAAATGTAGATGCAGCTGGCAGCTGGCAGCTGGCAGCTGGGAAAGCTGCGAGCACT 71967
Qy 1111 CAGTGCAGCCCTGTGAAAGCAGCAGGATGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1170
Db 71968 CAACTCCAACCCATGAGATCAGCCACATGGGCTAC--TCCCAGGGGAAGCCACAGAGGCA 72025
```



```
Db 63617 GCTCCAGAAAGCAAGCAATGCTTGGCAGTGTACCTGGTGTCTAATCTGCAGGG 63558
Qy 826 GCAAGAAGTCAAGAAATTTGGGAA---CCTCCAATCAGATTTTCAGAAATATAT 882
Db 63557 ATACAGAATGCAAGAGTTGTAGGCCACTATGGCTTCACCTAGATTCAGAGGATGAT 63498
Qy 883 GGAACCCCTGGATGCCAGCAGCAAGTTGCTCTAGGGTGGGCTCTCATGAGAGACC 942
Db 63497 CTAAGAAGATGGGACCCAGCAGAGCTTGTACAGAGGAGCAGCCATGGCTGAGAGCC 63438
Qy 943 TC 944
Db 63437 CC 63436

RESULT 10
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match 3.5%; Score 48.8; DB 1; Length 7218;
Best Local Similarity 4.9%; Pred. NO. 5.7e-05;
Matches 17; Conservative 192; Mismatches 139; Indels 0; Gaps 0;

Qy 358 TGGCTAATGAGGAACAAATTTGGCTTAATAAGGAAGAAATTTCTAAGCAGCAAGCATT 417
Db 1462 TAGTTAAAGATAGAAGAAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1403
```

```
Qy 418 CAAGAGGTGACTTGGTGCTGTTAAAGGCATTTCAGTTTCATAAGGAGGCAGACATAG 477
Db 1402 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1343
Qy 478 AGTTTCAGAAAATTTGCACCTTGACAATGTGATAAAAGAAAACCCATTTTCTCAGGG 537
Db 1342 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1283
Qy 538 AAATTCAGCTGGCTGCAGAAATTTGCATATGTAATGAGGAGCTGAATGTTAATCCTCAA 597
Db 1282 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1223
Qy 598 GACAATGGGAAATATCTCTCGACATGTGCAGAGTCTTCACAGCAGTCCATCAATCA 657
Db 1222 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1163
Qy 658 CTGGCTGGAGGCTAGGAGAAATGGTTTGTGGAGCAGCCAGGG 705
Db 1162 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1115

RESULT 11
US-08-463-911-6
; Sequence 6, Application US/08463911
; Patent No. 5869330
; GENERAL INFORMATION:
; APPLICANT: Scherer, Philipp E.
; APPLICANT: Lodish, Harvey F.
; TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
; TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,911
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WH195-05
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 73...804
US-08-463-911-6

Query Match 3.5%; Score 48.4; DB 2; Length 1313;
Best Local Similarity 48.8%; Pred. NO. 2.7e-05;
Matches 164; Conservative 0; Mismatches 166; Indels 6; Gaps 1;

Qy 41 ATGTAGAAATGCTGGACCTCCAGCACACCCCGCCCGCCAGAGAAAGTGGGCTCTG 100
Db 281 ATCCAGGTCTTATTGGTCTTAAGGAGAGATCGGTGAACCGGAGTACCCGGGCTGAAG 340
```

```

QY    101 GTGCACCAGGTTTACCAACAATATACAGGGAGAATAAGTGAAATGACAAAATGCCCTGTC   160
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    341 GTC CCGAGGCTTCCGGGAATCCAAGCAGGAAGAGAACCTGGAGAAAGGTGCCT---   397
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY    161 CTGATATAGAAAGGTCAGCCTTTACTGTGAAGCTCAGTGGAAAACTTCCTCTTCCTTTCA   220
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    398 ---ATGTATTACCGCTCAGCATTCAGTGTGGGATTTGGAGACTTACGTTACTATCCCACA   454
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY    221 AGCCCATCATCTTCACAGGGTGCTGTACANCCCCAGAGGGAATTAAAGGAGGCCATGG   280
      ||||| |||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    455 TGCCCATTCGCTTTACCAAGATCTTACAATCAGCAAACCACTATGATGGCTCCACTG   514
      ||||| |||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY    281 GAGTCCTTTGCTTCAGGGTGCCCTGGGAATTACTACTCCAGCTTTTGATGTTGAGCTGCATC   340
      | ||| |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||
Db    515 GTAATTCCACTGCAACATCTCTGGCTGTACTACTTTCCCTACCACTCACAGTCTATA   574
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY    341 ATTGCAAGGTGAATATTGTCCTAATCAGGAAGCAAA   376
      ||||| |||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    575 TGAAGGATGTGAAGTCAGCCTCTTCAAGAAGGACA   610
      ||||| |||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-09-140-804-9
; Sequence 9, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Humes, Jacqueline M.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49
; CURRENT APPLICATION NUMBER: US/09/140,804
; CURRENT FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: 60/056,983
; EARLIER FILING DATE: 1997-08-26
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows version 3.0
; SEQ ID NO 9
; LENGTH: 4517
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-140-804-9
```

```

US-09-188-930-217
; Sequence 217, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 217
; LENGTH: 1107
; TYPE: DNA
; ORGANISM: Rat
US-09-188-930-217

Query Match          3.1%; Score 42.4; DB 3; Length 1107;
Best Local Similarity 50.5%; Pred. No. 0.0021;
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps

Qy 168 AGAAGTCAAGCCTTTACTCTGAAGCTCAGTGGAAACTTCCTCTTCCTTTCAAGGCCAT 227
    || || || || || || || || || || || || || || || || || || || || ||
Db 614 AGCCAAGTCGGCTTTTCGCTGGCGGTAAACCAAGAGTTACCCACGTGAGCGACTGCCCAT 673

Qy 228 CATCTTCACAGGGTCTGTGTACAAATGCCAGAGGGATTTAAAGGAGCCATGGGAGTCTT 287
    || || || || || || || || || || || || || || || || || || || || ||
Db 674 CAAGTTTGACAAGATTCGTGATGAATGAGGGAGGCCACTACAATGCATCCAGTGGCAAGTT 733

Qy 288 TGCTTCGAGGGTGCTCGGGAATTACTACTCCAGCTTTTGATGTTGAGCTGCATCATGCAAA 347
    || || || || || || || || || || || || || || || || || || || || ||
Db 734 CGTCTGACGGTGCAGGGGATCTATTACTTTTACCTATGACATTACCTGGCCACACAACA 793

Qy 348 GGTGAATATTTGGCTAATAGGAA 371
    || || || || || || || || || || || || || || || || || || || || ||
Db 794 CCTGGCCATCGGCTAGTGCACAA 817
    || || || || || || || || || || || || || || || || || || || || ||

```

RESULT 13

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1276 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 46..786
US-08-463-911-1

Query Match          2.9%; Score 40; DB 2; Length 1276;
Best Local Similarity 49.1%; Pred. No. 0.013;
Matches 106; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

Qy 161 CTGATATAGAAAGTCAAGCTTTTACTGTGAAGCTCAGTGGAAACCTTCCTCTTCTTTCA 220
   ||||| | | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 377 CTTATATGATCGCTCAGCGTTCAGTGTGGGCTGGAGACCCGGCTCACTGTCCCAATG 436
   ||||| | | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Qy 221 AGCCATCATCTTCACAGGGGTCTGTACATGCCCAGAGGGATTTAAAGGAGGCCATGG 280
   ||||| | | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 437 TACCCATTCGCTTTACTAAGATCTTCTACAAACCAACAATCATATGACGGCAGCACTG 496
   ||||| | | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Qy 281 GAGTCTTTGCTTCAGGGGTGCTGGGAATTACTCTCCAGCTTTGATGCTTGCGTGCATC 340
   ||||| | | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 497 GCAAGTTCTACTGCAACATTCGGGACTCTACTCTCTTACCACATCACTCGGTGTACA 556
   ||||| | | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Qy 341 ATTCGAAGGTGAATATTTGGCTAATGAGGAAGCAAA 376
   ||||| | | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 557 TGAAGATGTGAAGGTGAGCCTCTTCAAGAAGGACA 592
   ||||| | | |||| | |||| | |||| | |||| | |||| | |||| | |||| |

RESULT 15
US-08-543-246B-8
; Sequence 8, Application US/08543246B
; Patent No. 6262244
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: DNA and amino acid sequence specific for
; TITLE OF INVENTION: natural killer cells
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Michael W. Glynn
; ADDRESSEE: No. 6262244artis Corporation
; STREET: 564 Morris Avenue
; CITY: Summit,
; STATE: NJ
; COUNTRY: US
; ZIP: 07901-1027
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/543,246B
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/676,663
; FILING DATE: 28-MAR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/02469
; FILING DATE: 27-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/122,514
; FILING DATE: 24-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Kassenoff, Melvyn M.
; REGISTRATION NUMBER: 26,389
```

```
; REFERENCE/DOCKET NUMBER: 118-7704/PCT/CONT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-522-6927
; TELEFAX: 908-522-6955
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1755 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 339..986
; NAME/KEY: mat_peptide
; LOCATION: 339..986
US-08-543-246B-8

Query Match          2.8%; Score 38; DB 4; Length 1755;
Best Local Similarity 62.8%; Pred. No. 0.07;
Matches 59; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 1283 TTCGACATTATATGGGCGCCGTACCCCTTTTGTGTTGGCAATTTTTTCATTTGGAATG 1342
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 1564 TTTGTACTTGCTTGAGTCCCATGACTGTTCTCTTCCTCTCTTCTCTCTTTTGAATAG 1623
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Qy 1343 CCGTATTTACCAATGCGCTGACCTCCATTTGAT 1376
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 1624 TAATATCCATCCTATGTTTGTCCCACTATTGTAT 1657
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||

Search completed: February 19, 2003, 23:16:17
Job time : 910.904 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 15:08:07 ; Search time 476.571 Seconds  
(without alignments)  
6506.903 Million cell updates/sec

Title: US-09-997-610-1-copy\_2\_1378

Perfect score: 1377

Sequence: 1 atagtggtacatacctgtctt.....gcctgtacctcattgtatg 1377

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 segs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1281	93.0	1338	24	AA144066 Human genset metab
2	663	48.1	2590	22	AAH14327 Human cDNA sequenc
3	643.8	46.8	109201	24	ABQ88125 Human osteoblast d
4	640	46.5	155074	24	ABN85735 Human genomic regi
5	639.8	46.5	3093	22	AA503687 Rhesus gene locus:
6	639.8	46.5	9236	22	AA503689 Rhesus gene locus:
7	639.8	46.5	9236	22	AA503690 Rhesus gene locus:
8	639.8	46.5	9241	22	AA503688 Rhesus gene locus:
9	631	45.8	98690	24	ABK12169 Human DNA represen

10	618.4	44.9	6063	22	AA103635 Human reproductive
11	618.2	44.9	465237	24	ABQ87681 Human oestrogen re
12	618.2	44.9	465237	24	ABQ87681 Human oestrogen re
13	617.4	44.8	3203	22	AA051134 Human secreted pro
14	616.8	44.8	6063	22	AA103634 Human reproductive
15	602.8	43.8	1736	22	AAH13678 Human cDNA sequenc
16	595.6	43.3	22509	22	AA544505 Human LEKTI DNA cl
17	595.6	43.3	31529	22	AA544505 Human LEKTI DNA cl
18	590.8	42.9	1278	22	AA051178 Human secreted pro
19	590	42.8	32127	22	AAI99255 Human excretory re
20	590	42.8	32127	22	AAI99255 Human kidney relat
21	590	42.8	72215	22	AAI63605 Human immune/haema
22	580.4	42.1	1485	23	AAK86832 DNA encoding novel
23	563.8	40.9	8387	22	ABAI4892 Human nervous syst
24	563.8	40.9	8387	22	ABAI4892 Human immune/haema
25	559.2	40.6	7726	22	ABA21079 Human nervous syst
26	558.2	40.5	1580	23	AA87262 DNA encoding novel
27	557	40.5	159400	24	ABQ88126 Human osteoblast d
28	556.6	40.4	1621	23	AA567924 DNA encoding novel
29	555.8	40.4	81369	21	AA57997 Human T gene DNA.
30	551.6	40.1	1306	23	AA564678 DNA encoding novel
31	549	39.9	3049	22	AAH16637 Human cDNA sequenc
32	543	39.4	1405	22	AAK70265 Human immune/haema
33	539.8	39.2	1405	22	AAK70265 Human immune/haema
34	523	38.0	1946	22	ABAI4227 Human breast cell
35	523	38.0	1946	22	ABAI4227 Human foetal liver
36	523	38.0	1946	22	ABA24459 Probe #2925 for ge
37	523	38.0	1946	22	AAK02964 Human brain expres
38	523	38.0	1946	22	AAK28410 Human bone marrow
39	523	38.0	1946	22	AAI12975 Probe #2908 for ge
40	523	38.0	1946	22	AAI34334 Probe #3020 used t
41	523	38.0	1946	22	AAI02893 Probe #2884 used t
42	523	38.0	1946	24	ABS02918 Human genome-deriv
43	519.4	37.7	22756	22	AA540321 DNA encoding human
44	519.4	37.7	22756	22	AA540321 DNA encoding human
45	515.2	37.4	923	23	AA566623 DNA encoding novel

ALIGNMENTS

RESULT 1  
AA144066

ID AA144066 standard; cDNA: 1338 BP.

XX AA144066;

AC AA144066;

XX 27-SEP-2002 (first entry)

DT Human genset metabolic gene (GMG-9) cDNA sequence.

DE Human; gene; ss: gene therapy; genset metabolic gene; GMG-7A; GMG-7B;  
KW GMG-8; GMG-9; GMG-10; GMG-11; metabolic-related disorder; obesity;  
KW impaired glucose tolerance; insulin resistance; Syndrome X;  
KW Type II diabetes; hyperlipidaemia; atherosclerosis; hypertension;  
KW heart disease; cardiac insufficiency; coronary insufficiency;  
KW high blood pressure; insulin sensitiser;  
KW non-insulin dependent diabetes mellitus.

XX Homo sapiens.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 1..1338

FT /\*tag= a

FT /product= "Human GMG-9 protein"

FT /note= "No stop codon is given"

XX WO200255694-A2.

PN 18-JUL-2002.

XX 15-JAN-2002; 2002WO-IB01215.

XX

XX

XX

XX

XX

XX

XX

XX

XX 16-JAN-2001; 2001US-262235P.  
PR (GEST ) GENSET.  
XX  
PA Erickson MR, Bour BA, Bihain B, Tanaka H;  
XX  
DR WPI; 2002-557821/59.  
DR P-PSDB; AAO15423.  
XX  
XX Treating or preventing a metabolic-related disease or disorder, e.g.  
PT obesity, impaired glucose tolerance, insulin resistance, Syndrome X, or  
PT Type II diabetes, comprises administering Genset Metabolic Genes -  
XX  
PS Disclosure; Page 122-124; 128pp; English.  
XX  
CC The invention comprises the amino acid and coding sequences of six human  
CC genset metabolic genes (GMG-7A, GMG-7B, GMG-8, GMG-9, GMG-10 and GMG-11).  
CC The GMG DNA and protein sequences of the invention are useful for  
CC treating or preventing metabolic-related disorders, such as: obesity;  
CC impaired glucose tolerance; insulin resistance; Syndrome X; Type II  
CC diabetes; hyperlipidaemia; atherosclerosis; hypertension; and heart  
CC diseases (e.g. cardiac insufficiency, coronary insufficiency or high  
CC blood pressure). The GMG DNA and protein sequences of the invention may  
CC also be used as insulin sensitizers - for improving insulin sensitivity  
CC in persons with non-insulin dependent diabetes mellitus. The present cDNA  
CC sequence encodes the human GMG-9 protein.  
XX  
SQ Sequence 1338 BP; 359 A; 288 C; 350 G; 341 T; 0 other;

Query Match 93.0%; Score 1281; DB 24; Length 1338;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 1284; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 89 TGGGGCTCTCGTGCACACAGGTTTACACAAATATACAGGAGAAATAGTGAATGACAA 148  
DB III III III III III III III III III III III III III III III III III  
50 TGGAAATGCTGTGCTGCCCGAGGTTTACCACAAATATACAGGAGAAATAGTGAATGACAA 109  
QY 149 AATGCCCTGTCTGATAGAAAGTCAGCCCTTACTGTGAAGCTCAGTGGCAAACTTC 208  
DB III III III III III III III III III III III III III III III III III  
110 AATGCCCTGTCTGATAGAAAGTCAGCCCTTACTGTGAAGCTCAGTGGCAAACTTC 169  
QY 209 CTCCTTCCTTTCAAGCCCATCATCTTACAGGGGTGCTGTACAAATGCCAGAGGGATTTAA 268  
DB III III III III III III III III III III III III III III III III III  
170 CTCCTTCCTTTCAAGCCCATCATCTTACAGGGGTGCTGTACAAATGCCAGAGGGATTTAA 229  
QY 269 AGGAGCCATGGAGTCTTTGGTTCGAGGGTGCCTGGGAATTTACTTCCAGCTTTGATG 328  
DB III III III III III III III III III III III III III III III III III  
230 AGGAGCCATGGAGTCTTTGGTTCGAGGGTGCCTGGGAATTTACTTCCAGCTTTGATG 289  
QY 329 TTGAGCTGCATCATTTGCAAGGTGAATATTTGGCTAATGAGGAAGCAAAATTTTGGCTAATA 388  
DB III III III III III III III III III III III III III III III III III  
290 TTGAGCTGCATCATTTGCAAGGTGAATATTTGGCTAATGAGGAAGCAAAATTTTGGCTAATA 349  
QY 389 AGGAAGAAATTTCTAAGCAGCAAAAGCAATTCAGAGGTGACTTTGGGTGCTGTTTAAAGGCAT 448  
DB III III III III III III III III III III III III III III III III III  
350 AGGAAGAAATTTCTAAGCAGCAAAAGCAATTCAGAGGTGACTTTGGGTGCTGTTTAAAGGCAT 409  
QY 449 TCAGTTTTCATAAGGAGGAGCAGCATAAGAGTTTCAGAAATTTGCACCCCTGACAAATGTGA 508  
DB III III III III III III III III III III III III III III III III III  
410 TCAGTTTTCATAAGGAGGAGCAGCATAAGAGTTTCAGAAATTTGCACCCCTGACAAATGTGA 469  
QY 509 TAAAGAAAGAAACCCATTTTCTGAGGGGAATTTCAAGCTGGCTGCAGAAATTTGCATAT 568  
DB III III III III III III III III III III III III III III III III III  
470 TAAAGAAAGAAACCCATTTTCTGAGGGGAATTTCAAGCTGGCTGCAGAAATTTGCATAT 529  
QY 569 GTAATGAGAGCTGAATGTTAATCTCTCAAGACAATGGGGGAAATATCTCTGGACATGTC 628  
DB III III III III III III III III III III III III III III III III III  
530 GTAATGAGAGCTGAATGTTAATCTCTCAAGACAATGGGGGAAATATCTCTGGACATGTC 589  
QY 629 AGAGGTCTTCACAGCAGTCCATCAATCACTGGCCCTGGAGGCCCTAGGAGAAAATGGTTTT 688  
DB III III III III III III III III III III III III III III III III III  
590 AGAGGTCTTCACAGCAGTCCATCAATCACTGGCCCTGGAGGCCCTAGGAGAAAATGGTTTT 649

QY 689 GTGGGACAGGCCAGGGTCCCTGTGTGTGCAGCCTAGAGACTTGTGTGCCCTGTGTGTC 748  
DB III III III III III III III III III III III III III III III III III  
650 GTGGGACAGGCCAGGGTCCCTGTGTGTGCAGCCTAGAGACTTGTGTGCCCTGTGTGTC 709  
QY 749 CAGTTTAATTTACGCTGTGGCTTCAGAGGGTCAAGCCCCAAGCCTTGGCAGCTTCCAAAGTG 808  
DB III III III III III III III III III III III III III III III III III  
710 CAGTTTAATTTACGCTGTGGCTTCAGAGGGTCAAGCCCCAAGCCTTGGCAGCTTCCAAAGTG 769  
QY 809 GTGTTGAGCCTGTGGTGCACAAAGAGTCAAGAAATTTGAGGTTTGGGAACTCCAAATCAGAT 868  
DB III III III III III III III III III III III III III III III III III  
770 GTGTTGAGCCTGTGGTGCACAAAGAGTCAAGAAATTTGAGGTTTGGGAACTCCAAATCAGAT 829  
QY 869 TTCAGAAGATATATGGAACCCCTTGGATGCCAGGCAGAGATTTTGTGTAGGGGTGGGT 928  
DB III III III III III III III III III III III III III III III III III  
830 TTCAGAAGATATATGGAACCCCTTGGATGCCAGGCAGAGATTTTGTGTAGGGGTGGGT 889  
QY 929 CCTCATGGAACACCTCTGCAAGGGTAGTACAAAGGAAATTTGGGTGGGAGCCCCCAC 988  
DB III III III III III III III III III III III III III III III III III  
890 CCTCATGGAACACCTCTGCAAGGGTAGTACAAAGGAAATTTGGGTGGGAGCCCCCAC 949  
QY 989 ACAGAGTCCCCAGTGGGGTCCATCTAGTAGAGCTGTGAGAAGAGTCCACCATCCTCCA 1048  
DB III III III III III III III III III III III III III III III III III  
950 ACAGAGTCCCCAGTGGGGTCCATCTAGTAGAGCTGTGAGAAGAGTCCACCATCCTCCA 1009  
QY 1049 GACTCCAGAAGGTTAGATCCACTGACAGCTTGCAGCATGTGCTGAAAAATCCACAGACA 1108  
DB III III III III III III III III III III III III III III III III III  
1010 GACTCCAGAAGGTTAGATCCACTGACAGCTTGCAGCATGTGCTGAAAAATCCACAGACA 1069  
QY 1109 CTCAGTGCCAGCTGTGAAAAGCAGCAGGGATGGAGTCTTACCCCTACAAAAACCGTAGTGG 1168  
DB III III III III III III III III III III III III III III III III III  
1070 CTCAGTGCCAGCTGTGAAAAGCAGCAGGGATGGAGTCTTACCCCTACAAAAACCGTAGTGG 1129  
QY 1169 CAGAGCTGACCAAGACCGTGGGAATCTACTCTTGCATTTGTATGACCTTGGAGCTGAGAC 1228  
DB III III III III III III III III III III III III III III III III III  
1130 CAGAGCTGACCAAGACCGTGGGAATCTACTCTTGCATTTGCATGACCTGGAGCTGAGAC 1189  
QY 1229 ATGGAGTCAAAAAGAGATCATTTTGGAGCTTTAAGATTTTCACTGCCCACTGGATTTTCGGA 1288  
DB III III III III III III III III III III III III III III III III III  
1190 ATGGAGTCAAAAAGAGATCATTTTGGAGCTTTAAGATTTTCACTGCCCACTGGATTTTCGGA 1249  
QY 1289 CTTATATGGGGCCGTACCCCTTTGTTTTGGCCAAATTTTTCATTTGGAAGTGGCGTAT 1348  
DB III III III III III III III III III III III III III III III III III  
1250 CTTATATGGGGCCGTACCCCTTTGTTTTGGCCAAATTTTTCATTTGGAAGTGGCGTAT 1309  
QY 1349 TTACCCAATGCTGTACCTTCCATTTGTATG 1377  
DB III III III III III III III III III III III III III III III III III  
1310 TTACCCAATGCTGTACCTTCCATTTGTATG 1338

RESULT 2  
AAH14327  
ID AAH14327 standard; cDNA; 2590 BP.  
XX  
AC AAH14327;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA sequence SEQ ID NO:11697.  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.  
XX 09-JUN-2000; 2000JP-0241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Mishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
XX full-length cDNAs -  
PS Claim 8; SEQ ID 11697; 2537pp + CD ROM; English.  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
XX of the present invention.  
SQ Sequence 2590 BP; 722 A; 526 C; 670 G; 672 T; 0 other;  
Query Match 48.18; Score 663; DB 22; Length 2590;  
Best Local Similarity 80.78; Pred. No. 7.5e-197;  
Matches 826; Conservative 0; Mismatches 160; Indels 37; Gaps 3;  
QY 391 GAAGAAATTTCTAGCAGCAGAGCATTCAGAGGTTGACTTGGTGCTCTTAAGGCATTC 450  
DB 609 GAAGAAATTTCTAGCAGCAGAGCATTCAGAGGTTGACTTGGTGCTCTTAAGGCATTC 668  
QY 451 AGTTTCATAGGGAGGAGCAGACATAAGAGTTTCAGAAAAATTTGCACCCCTGACAATGTGATA 510  
DB 669 AGTTTCATAGGGAGGAGCAGACATAAGAGTTTCAGAAAAATTTGTAGCCTGACTATGCATATA 728  
QY 511 AAAAGAAAAACCCATTTTCTGAGGGGAAATTCAGCTGGCTGCGCAGAAAAATTTGCATATGT 570  
DB 729 GAAAGAAAAACCCAGTTTCTGCTGGGAGAAATTCAGGCAGCTGCGCAGAAATTCGCATAAGC 788  
QY 571 AATGAGGAGCTGAATGTTAATCCCTCAAGACAAATGGGAAAAATATCTCTGGACATGTCTAG 630  
DB 789 ATCAAGAAAGCCCTAATGTTTAAATCCCGAGGACCATGGGAAAAATGTCACAGGCCATGTCTAG 848  
QY 631 AGGCTTCCACAGCAGCTCCATCAATCACTGCGCTGGAGGCTCTAGGAG--AAAATGGTTT 687  
DB 849 AGACCTTCACAGCAGCCCTCTGTCTACAGGCCCCAGAGATCCAGGAGGAAAAATGGTTT 908  
QY 688 TGTGGACAGGCCAGGCTCCCTGT 747  
DB 909 TATGGCCAGGACCAAGGCTCCCTGT 968  
QY 748 CCAGTTAATTC-----GCTGTGCTTCAGAG 774  
DB 969 CCAGCTGTCCAGCCATGGCTGAAAGGGGCCAATGTAGAGCTCATGTGTGTGTGTGTGTGTGTGT 1028

QY 775 GGTGCAAGCCCCAAGCCTTGGCAGCTTCCAACTGGTGTGTTGAGCCTGTGGTGCACAAGAG 834  
DB 1029 GGTGCAAGCCCCAAGCCTTGGCAGCTTCCAACTGGTGTGTTGAGCCTGTAGGTACACAGAAG 1088  
QY 835 TCAAGAAATTTGAGTTTGGGAACCTCCAACTCAGATTTTCAGAAATATATATGAAACCCCTGG 894  
DB 1089 TCAAGAAATTTGAGTTTGGGAACCTCCAACTCAGATTTTCAGAAATATATATGAAATGCTCTG 1148  
QY 895 ATGCCCAGCAGAAAGTTTGTCTAGGGTGGGGTCTCTCATGTGAGAAACCTCTCTGCAAGGGTA 954  
DB 1149 ATGTCCAGCAAAAGTTTGTCTAGTGGCAGGGCCCTCATGTGAGAAACCTCTCTGCTAGGGCA 1208  
QY 955 GTACAAAAGGGAATTTGGGTGGGAGCCCCACACAGAGTCCCGAGTGGGGTCCCATCT 1014  
DB 1209 ATGCAGAAAGGAATTTGGGGCCGGAGCTCCACAGTAGAGTCCCTAATGGGGCATTATCT 1268  
QY 1015 AGTAGAGCTGTGAGAAAGTCCACCATCTCCAGACTCCAGAAAGGTAGATCCACTGAC 1074  
DB 1269 AGTGGAGCTGTGAGAAAGGAGGCCACCGTCTTCAGACCCCGCATGGTGGATCCACTGAC 1328  
QY 1075 AGCTTGCAGCATGTGCTGAAAAATCCACAGACACTCAGTGCAGCTCTGTAAGCAGCA 1134  
DB 1329 AGCTTGCACCATGCGCTGGAAAAAGCCGAGACACTCAATGCGAGTCCGTGAAAGCAGCC 1388  
QY 1135 GGGATGGAGTGTGTACCTTACAAACCCGTAGTGGCAGAGCTGACCAAGACCCTGGGAATC 1194  
DB 1389 AGGAGGAGGAGCTGTACCTTGCAAAGCCACAGGGGCGAAGAACTCCAAAGGCCATGGGTCCC 1448  
QY 1195 TACCTTCCTGATTCATGACCTGGACGTGACAGCATGGAGTCAAAAGAGATCATTTTGA 1254  
DB 1449 CACCTTTTGCATFCAGCGTGAACCTGAATGTGAGACCTAGAGTCAAAAGGAGATCATTTTGA 1508  
QY 1255 GCTTTAAGATTTGACTGCCCCACTGGATTTGCGACTTATATGGGCCCCGTA-CCCCCTTTG 1313  
DB 1509 GCTTTAAGATTTGATGTTGCTCTGCTGATTTTGGACTTGCATGGTCTCGACGCCCTTTG 1568  
QY 1314 TTTTGGCCAAATTTTTCATTTTGGAACTGCGCTATTTACCCAAATGCTGTACTCCATTTG 1373  
DB 1569 TTTTGGCCAAATGCTCTCCCATTTTGGAAATGCTGTATTTACCCAAATACCTGTACCCCTATCG 1628  
QY 1374 TAT 1376  
DB 1629 TAT 1631  
RESULT 3  
AB088125/c  
ID AB088125 standard; cDNA; 109201 BP.  
XX AC AB088125;  
XX DT 18-SEP-2002 (first entry)  
XX DE Human osteoblast differentiation related cDNA SEQ ID NO 32.  
XX DE Human; osteoblast; stem cell differentiation; bone tissue deposition;  
XX KW osteoporosis; osteopathic; ss.  
XX OS Homo sapiens.  
XX PN WO200250301-A2.  
XX PD 27-JUN-2002.  
XX PF 18-DEC-2001; 2001WO-US48276.  
XX PR 18-DEC-2000; 2000US-255882P.  
XX PR 24-APR-2001; 2001US-285691P.  
XX XX (GENE-) GENE LOGIC INC.  
XX PA (PROC ) PROCTER & GAMBLE CO.  
XX XX

PI Ji D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;  
PI Mertz L;  
XX  
DR WPI; 2002-557663/59.  
XX  
PT use of genes and their expression profiles associated with osteoblast  
PT differentiation for screening modulators bone formation, for diagnosing  
PT or treating e.g. osteoporosis, or as markers for the differentiation  
PT process  
XX  
XX  
PS Claim 1; SEQ ID NO 32; 78pp + Sequence Listing; English.  
XX  
CC The invention relates to genes and their expression profiles are used  
CC for:  
CC (a) screening modulators of precursor stem cell differentiation into  
CC osteoblasts, or bone tissue deposition;  
CC (b) diagnosing abnormal deposition of bone tissue, abnormal rate of  
CC osteoblast formation or osteoporosis; or  
CC (c) treating or monitoring treatment of the conditions cited in (b), or  
CC monitoring the progression of bone tissue deposition.  
CC Specific conditions include postmenopausal osteoporosis, glucocorticoid  
CC osteoporosis or male osteoporosis, osteopenia, osteodystrophy,  
CC drug-induced abnormalities in bone formation or bone loss, conditions  
CC that involve altered bone metabolism (e.g. idiopathic juvenile  
CC osteoporosis), skeletal disease linked to breast cancer, mastocytosis,  
CC Fanconi syndrome or fibrous dysplasia. The present sequence is that of an  
CC osteoblast differentiation associated cDNA marker of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 109201 BP; 32871 A; 23488 C; 22108 G; 30734 T; 0 other;  
Query Match 46.8%; Score 643.8; DB 24; Length 109201;  
Best Local Similarity 79.3%; Pred. No. 7.3e-190;  
Matches 843; Conservative 0; Mismatches 177; Indels 43; Gaps 5;  
QY 351 GAATATTTGGCTAATGAGGAACAATTTTGGCTAATAGGAAGAAATTTCTAAGCAGCA 410  
DB 22659 GAACTTTGAACCTTGAGAGAGATGATTTAGGGTATCTGGCAGAGAATAATTTCTAAGCAGCA 22600  
QY 411 AAGCATTTAAGAGGTGACTTTGGGTGCTGTAAAGGCAATTCAGTTTCATTAAGGGAGGCAGA 470  
DB 22599 AAGCATTTAAGAGATGACTTTGGGTGCTGTAAAGGCAATTTATTTTATAAGGGAGAGCAGA 22540  
QY 471 GCATAAGAGCTTCAGAAAATTTTCACCCCTGCACAAATGTGATAAAAAGAAAACCCATTTTC 530  
DB 22539 GCATAAAGTTTGGAAAATTTGCATCTCTGACAAATGTGATAGAAAAGACAATCCTATTTTC 22480  
QY 531 TGAGGGGAAATTCAGCTGGCTGCAGAAATTTGCATATGTAATGAGGAGCTGAATGTTAA 590  
DB 22479 TGAGGAGACATTCAGTCAAGCTGCAGAAATTTGCATAAGTAATGAGGAGCTTAATTTAA 22420  
QY 591 TCCTCAAGACATGGGAAAATATCTCTCGACATGTCTAGAGGTCTTCACACAGTCCAT 650  
DB 22419 TCCCAAGACATGGGAAAATGTCTCCAGGCAATGTCTCAGAAAGTCTTCACAGACGCCCT 22360  
QY 651 CAATTCACGTGGCTGGAGGCCTA---GGAGAAAATGGTTTCTGGGACAGGCCAGGGTC 707  
DB 22359 CCCATCAGAGCCTGGAAGCCTTAGAGGAAAAGTGTTTTTTGGCCAGGCCACAGGGTC 22300  
QY 708 CCTGTGCTGTGTGCAGCCTTAGAGACTTGTGTGCCCTGTGTCCAGTGAATTCAGCTGTGGC 767  
DB 22299 CTCATGCTGTGTGCAGCCTTAGAGACTTGTGTGCCCTGCATCCCGCAGCTGCAGCTATGGG 22240  
QY 768 T-----TCAGAGGTGCAAGCCCCAAGCCCTT 793  
DB 22239 TGAAGAGGGCCCAACACAGAGCTCAACAACAGAGAGTCTAGAGGTGCAAGCCTCAAGCTTT 22180  
QY 794 GCAGCTTCCAAGTGGTGTAGCCTGTGGGTGCAAGAGTCAACAATTTGAGTTTGGG 853  
DB 22179 GGTAGCTCCATGTGGTGTGATTCTGCGAGTGCACAGAAAGCAAGAAATTTGGGTTTGGG 22120

QY 854 AACCTCCAATCAGATTTTCAGAAATATATGAAACCCCTGGATGCCCGCAGGAAAGTGTG 913  
DB 22119 AACTTGTGCTAGATTC---AGATGTATGAAATCTGTTGGATGTCCAGGCAGAACTTTG 22063  
QY 914 CTGTAGGGTGGGGTCTCTCATGGAACCTCTGCAAGGTAGTACAAAAGGAAATGTTG 973  
DB 22062 CTATGAG---GGGGTCTCTCATGTAATAATCTGCTAGGCAAGTGCAGAAAGAAATGTGG 22005  
QY 974 GTGTGGAGCCCCACACAGAGTCCCGCAGTGGGCTCCATCTAGTAGAGCTGTGAGAAGAA 1033  
DB 22004 GTTAGAGAGCCCCCATACAGAGTCTCTACTGGGCACCACTAGTGGAGCTGTGAGAAAAG 21945  
QY 1034 GTCCACCAATCTCCAGACTCCAGAAGGTAGATCCACTGCACAGCTTGCAGCATGTGCCCTG 1093  
DB 21944 GGCCACCAATCTCCAGACCCCAAGATGGTAGTCCACTCACAGCTTGCATATATGTGCCCTG 21885  
QY 1094 AAAATCCACACACACTCAGTCCAGCCTGTGAAGCAGCAGGAGTGGAGTCTGTACCCT 1153  
DB 21884 GAAAAGCCATAGACACTCAACACCAAGGCTGTGAAAGCAGCCAGGAGGAGGCTGTACCCT 21825  
QY 1154 ACAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTCTTTCATTTGTCATG 1213  
DB 21824 GCAAGCCACAGGGTGAAGCTGACCAAGACTATGGGAACCTACTTCTTGCATCAGTGTG 21765  
QY 1214 ACCTGAGCTGAGACATGAGTCAAAAAGAGATCAATTTTGGAGCTTTAAGATTTGACTGCC 1273  
DB 21764 ACCTGAGTGTGAGACATGCTGTCAAAAGGAGATCCTTTTGGAGCTTTAAGATTTGACTGCC 21705  
QY 1274 CCACGTGATTTCCGACTTATATGGGCGCGTA---CCCTTTGTTTGGCCCAATTTTCCA 1332  
DB 21704 CTGCTAGATTTTCAGACTTGAATGGGCCCTGTAGCCCCCTTTGTTTGGCCCAATTTTCCA 21645  
QY 1333 TTTGGAACCTGCGTATTTTACCCTGCTGCTACCTCCATTTGTA 1375  
DB 21644 TTTGGAATGGCTGATTTACTCAATGCCCTGCATCCCTATTTGTA 21602  
RESULT 4  
ABN85735/C  
ID ABN85735 standard; DNA; 155074 BP.  
XX  
AC ABN85735;  
XX  
DT 30-SEP-2002 (first entry)  
XX  
DE Human genomic region containing the ltrpc5 gene SEQ ID NO 6.  
XX  
DE Human; ltrpc5; taste; cell signalling; TC-ICS; food; pharmaceutical;  
KW taste cell-specific ion channel subunit; chromosome 11p15-5; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200254069-A1.  
XX  
PD 11-JUL-2002.  
XX  
PF 26-DEC-2001; 2001WO-US49808.  
XX  
PR 29-DEC-2000; 2000US-259379P.  
XX  
PR 21-DEC-2001; 2001US-0026188.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
XX  
XX Zuker CS, Zhang Y;  
XX  
DR WPI; 2002-583632/62.  
XX  
PT Identifying modulators of taste signalling in taste cells for use in  
PT food and pharmaceutical industries to customize and regulate taste, by  
PT determining effect of the compound on a taste cell-specific ion channel  
PT subunit  
XX  
PS Disclosure: Page 212-301; 306pp; English.

XX The invention relates to identifying (M1) a compound that modulates taste  
CC signalling in taste cells, by contacting the compound with a eukaryotic  
CC host cell or cell membrane which expresses a taste cell-specific ion  
CC channel subunit (TC-ICS), and determining a functional effect of the  
CC compound upon a transmembrane ion flux of a predetermined ion,  
CC identifying a compound that modulates taste signalling in taste cells.  
CC (M1) is useful for identifying a compound that modulates taste signalling  
CC in taste cells, for identifying a compound that binds to a taste cell  
CC specific ion channel subunit and for modulating taste signalling in taste  
CC cells of a mammal, in particular a human. Modulators identified by (M1)  
CC are used by the food and pharmaceutical industries to customize taste,  
CC e.g. as additives to food or medicine so that the food or medicine tastes  
CC different to the subject who ingests it. Bitter medicines can be made to  
CC taste less bitter and sweet substance can be enhanced. The modulators are  
CC useful for pharmacological and genetic modulation of taste signalling  
CC pathways. The taste modulators can be directly administered to mammalian  
CC subjects for modulation of taste in vivo. The present sequence is that of  
CC the human genomic region (Chromosome 11p15-5) containing the *lirpe5* gene  
CC of the invention.  
XX  
XX Sequence 155074 BP; 32795 A; 41240 C; 44622 G; 36402 T; 15 other;

Query Match 46.5%; Score 640; DB 24; Length 155074;  
Best Local Similarity 80.6%; Pred. No. 1.4e-188;  
Matches 825; Conservative 0; Mismatches 160; Indels 39; Gaps 5;

Qy 390 GGAAGAAATTTCTAAGCAGCAAGCATTCAGAGGTGACTTGGGTGCTTAAAGGCATT 449  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 94900 GGAAGAAATTTCTAAGCTGCAAGAGCTTTCAGAGGTGACTTGGGTGCTTAAAGGCATT 94841  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 450 CAGTTTCAAGAGGAGGAGCAGCATAAGAGTTCAGAAATTTGCACCCCTGCACAAATGTGAT 509  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 94840 CAGTTTCAAGAGGAGGAGCAGCAGAG-ATAAAGATTCAAGAAATTTGCAGTCTGAAATGTGAT 94782  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 510 AAAAAAGAAAACCCATTTCTGAGGGGAAATTCAGCTGGCTGCAGAAATTTGCATATG 569  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 94781 AGAAATTAAGAAACCCATTTCTGAGGAGAAAGTCAAGCAGCGTGCAGAAATTTGCATATG 94722  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 570 TAATGAGGAGCTGAATGTAATCTCAAGACAATGGGCAAAATATCTCTCGACATGCA 629  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 94721 TAACAGAAAGCCAAATGTAATCACTAAGACAATGGGAAATATGCTCCAGGGCATGCA 94662  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 630 GAGGTCTTTCACAGCAGTCCATCAATCACTGGCCCTGGAGGCTAGGAG-AAAATGGTT 686  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 94661 GAGACCTTAGGAGCAGGCCCCCTCCATCCAGGCCAGAGGCTAGGAGGCAAAATGTT 94602  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 687 TTGTGGGACAGCCAGGGTCCCTGTGCTGTGTGTCAGCCTAGAGACTTGGTGCCTGTGT 746  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 94601 TCCTGGACTGGGCCAGGGCCCTCCATGCTGTGTGAGCCTAGGAGCTTGGTGCCTGTGT 94542  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 747 CCCAG------TTAATTCAGCTGTGGCTTCAGA 773  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 94541 CCCAGCCCTCTAGCCATAGCTAAAGGTGTCAAGGTACAGCTGGGCTGTGGCTTCAGA 94482  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 774 GGGTCAAGCCCAAGCCTTGGCAGCTTCCAGTGGTGTGTGAGCCTGTGGGTGCAAGAA 833  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 94481 GGGTCAAGCCCAACTCTTGTGAGTTCATGTTGGTGTGTGAGCCTGTGGGTGCAAGAA 94422  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 834 GTCAGAAATTCAGGTTTGGGAACCTCCAAATCAGATTTTCAGAGATATATGAAGCCCTG 893  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 94421 GTCAGAAATTCAGGTTTGGGAAGCTCTGCCCTAGATTTTCAG-AGATGTATGGAATGCTG 94363  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 894 GATGCCAGGAGGAATTTGCTGTAGGGGTGGGTGCTTCATGGAGAACCTCTGCAAGGTT 953  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 94362 GATGTCCAGGAGGAATTTGCTGCAGGGGTGGGTGCTTCATGGAGAACCTCTGCTAGGCT 94303  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 954 AGTACAAAAGGGAATGTTGGTGGGAGCCGCCACACAGAGTCCCAAGTGGGGTCTCCATC 1013  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 94302 AGTGAGGAAGGGAAGTGGGGTTGGAGCCGCCCATGCAAGTCCCACTGGGCACTGCC 94243  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 1014 TAGTAGAGCTGTGAGAGGAATGCACCATCTCTCCAGATCTCAGAAAGGCTAGATCCACTGA 1073  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 94242 TAGTGAGCTGTGAGAAGAGGGCCCACTGTCTCCAGATCCCAAGATGATAGATCAATCAA 94183  
Qy 1074 CAGCTTGGCAGCATGTCCTGAAAAATCCACAGACACTCAGTGCCAGCCTGTGAAAGCAGC 1133  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 94182 CAGCTTGGCAGCATGTCCTGGAAGAGCTGCACACACTCAATCCAGCCTGTGAAAGCAGC 94123  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 1134 AGGGATGGAGCTGTACCTTACAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAAT 1193  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 94122 CAGGAGGAGGCTGTACCTTGCAAAGCCAGAGGGGCGAGAGCTGCTCAAGACTATGGGAAC 94063  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 1194 CTACCTCTTGCAATTCATGACCTGAGCTGAGACATGGAGTCAAAAGAGATCATTTTGG 1253  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 94062 CCACCTCTTGCAATTCATGACCTGAGATGAGACATGGAGTCAATGGAGATCATTTTGG 94003  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 1254 AGCTTTAAGATTTGACTGCCCCACTGGATTCGGACTTATATGGGCCCCGTA-CCCTTTT 1312  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 94002 AGCATTAAGATTTAATGCCCACTGGATTTGCAACTTGCATGGGCGCTTTAGCCCCCTTC 93943  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 1313 GTTTTGGCCAAATTTTTCATTTGGAACCTGCGGTATTTATCCCAATGCCCTGTACCTCCATT 1372  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 93942 GTTTTGGCCAAATTTACTCCCATTTCAAATGGGTGTATTTATCCCAAGGCTGTACCCTAAT 93883  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 1373 GTAT 1376  
|||||  
Db 93882 GTGT 93879  
|||||  
RESULT 5  
RAS03687/C  
ID AAS03687 standard; DNA; 3093 BP.  
XX AAS03687;  
AC AC  
XX AC  
XX AC  
DT 29-AUG-2001 (first entry)  
XX  
XX Rhesus gene locus: RHD gene deletion in Rh negative haplotypes.  
DE Rhesus box: RHD positive; sequence length polymorphism; SSP; RHD; SMP1;  
KW RHE; Rh negative; blood group typing; blood transfusion; antigen C;  
KW haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.  
XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FH primer\_bind 32..54  
FT /\*tag= a  
FT /note= "Binding site of primer rez7"  
FT primer\_bind complement (3034..3054)  
FT /\*tag= b  
FT /note= "Binding site of primer rnb31"  
XX  
XX WO200132702-A2.  
PN  
XX  
XX 10-MAY-2001.  
PD  
XX  
XX 31-OCT-2000; 2000WO-EP10745.  
PF  
XX  
XX 02-NOV-1999; 99EP-0121686.  
PR 31-MAY-2000; 2000EP-0111696.  
PR  
XX  
XX (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTEMBERG.  
PA  
XX  
XX Flegel WA, Wagner FF;  
PI  
XX  
XX WPT; 2001-291052/30.  
DR  
XX  
XX New nucleic acid molecular structure, useful for detection of common  
PT RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1  
PT and RHCE genes -  
XX  
XX Example 10; Fig 5; 135pp; English.  
PS  
XX  
XX The sequence represents the coding sequence of Rhesus gene locus:  
CC

CC RHD gene deletion in Rh negative haplotypes. The Rhesus genes  
CC locus comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the  
CC Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box  
CC and/or the downstream Rhesus box. The RHD and RHCE genes are located at  
CC chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the  
CC RHD deletion in the common RHD negative haplotypes. The sequence has  
CC been used to design primers which are useful for: (1) the specific  
CC detection of the common RHD positive haplotypes in D-negative  
CC individuals; (2) blood group typing; (3) determining whether a patient  
CC can be transfused with Rhd negative blood and whether blood is suitable  
CC for transfusion to patients who should not be exposed to antigen C; (4)  
CC assessing the risk of a Rhd negative mother of conceiving or carrying an  
CC Rhd positive foetus. Anti-D antibodies are useful for treating pregnant  
CC women who are Rhesus D negative, where the foetus is not homozygous for  
CC the RHD gene to treat or prevent haemolytic disease of the newborn.

XX  
SQ Sequence 3093 BP; 891 A; 754 C; 619 G; 829 T; 0 other;

Query Match 46.5%; Score 639.8; DB 22; Length 3093;  
Best Local Similarity 79.1%; Pred. NO. 1.6e-189;  
Matches 841; Conservative 0; Mismatches 177; Indels 45; Gaps 5;

QY 351 GAATATTGGCTTAATCAGAGCAAAATTTGGCTAATRAGGAAGAAATTTCTAAGCAGCA 410  
DB 1377 GAACCTTGAACCTCAGAGAGATGATTTAGAGATCTGCGAGGAGAAATTTCTAAGCAGCA 1318  
QY 411 AAGCATTCAGAGGTGACTTGGTGCTGTTAAAGGCATTCAGTTTCATAAAGGAGGAGCA 470  
DB 1317 ANGCAATTCAGAGGTGACTTGGTGCTGTTAAAGCATTCAGTTTCATAAAGGAGGAGCA 1258  
QY 471 GCATAGAGTTTCAGAAAATTTGCACCCCTGCACAAATGTGATAAAAAAGAAAAA-CCCATTT 528  
DB 1257 GCATAAAAGTTCCGAAAATTTGCAGCCTGCACAAATGTGATAAAAAAGAAAAATTTCCCATTT 1198  
QY 529 TCTGAGGGAAATTCAGCTGCTCCAGNAATTTGCATATGTAATCAGGAGCTGAATGTT 588  
DB 1197 TCTGAGGAGAAATTCAGCTGCTCCAGNAATTTGCATATGTAATCAGGAGCTGAATGTT 1139  
QY 589 AATCCTCAGACAAATGGGAAAATATCTCTGGACATGTCAGAGTCTTCACAGCAGTCC 648  
DB 1138 AATTCAGACAAATGGGAAAATATCTCTGGAGGATGTCAGAGTCTTCATGGCAACCC 1079  
QY 649 ATCAATCACTGCGCTGGAGCCT-AGGAGAAAATGGTTTTTGGGACAGGCCAGGGTC 707  
DB 1078 CTCCCATCAGAGTCCAGAGGTATCAGGAAAATGGTTTTTGGCCAGGCCGCGGGTC 1019  
QY 708 CTTGCTGCTGTGCACCTTAGAGACTTGGTGCCCTGTGTCCCGATTA-----754  
DB 1018 CTGATGCTGTGTGCACCTTAGGAGCTTGGTGCCCTGCATCCAGCCACTCCCAACCATGA 959  
QY 755 -----ATTGAGCTGTGGCTTCAGAGGGTGCAAGCCCCCAAGCCCTT 793  
DB 958 CTGACGGGAGCAAGGTAGAGCTTGGGCTGTAGCTTGGGGAGTGCAGCCCCCAAGCCCTT 899  
QY 794 GGCAGCTTCCAAAGTGTGTGAGCTGTGGTGCAAGAGTCAAGAAATTTGAGTTGGG 853  
DB 898 GACAGCTTCCATGTGTGTGAGTGTGAGTGCAGAGTCAAGAAATTTGGGGTTTGA 839  
QY 854 AACCTCCAATCAGATTCAGAGATATATGGAACCCCTGGATGCCAGGCAAGATTG 913  
DB 838 AACCTTCGCTAGATTAAGAGAGATGTGCGGAAATTCGCTGGATGCCAGTCAAGATTG 779  
QY 914 CTGTAGGGTGGGCTCATGGAACCTCTGCAAGGGTGTAGTCAAAAGGGAATTTG 973  
DB 778 CTGAGGAGCAGGCGCTCATGAGATCTCTGCCAGGCGATGCGAAGGGAATTTGG 719  
QY 974 GGTGGAGCCCCACACAGAGTCCCCAGTGGGGCTTCATCTAGTAGAGCTGTGAGAAGAA 1033  
DB 718 GGTGAGAGACCCACACAGAGTCCCTACTTGGGGCCACCACTAGTGGAGCTGTGAGAAG 659  
QY 1034 GTCCACCATCTCCAGATCCAGAGGGTAGATCCACTGACAGCTTCAGCATGTGCTG 1093  
DB 658 G-----TCCTCCAGACCCCAAGATTTGGTAGATCCACCGACAGCTTGCACCGTGTACCTG 606

QY 1094 AAAAAATCCACAGACACTCAGTCCAGCCCTGTGAAAGCAGCAGGATGGAGTGTACCCCT 1153  
DB 605 GAAAGCTGCACACACTCAATCCAGCCCATGAAGCAGCTGAGAGGAGGCTGTACCCCT 546  
QY 1154 ACAAAACCGCTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTCTTTCATGTCATG 1213  
DB 545 GCAAAGGTACAGGGGAGAGCTGCGCAAGACCATGGGAACCCACCCCTTCCATCTCGGTG 486  
QY 1214 ACCTGGAGCTGAGACATGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTTGACTGCC 1273  
DB 485 ACCTGGATGTGAGATGTGGAGTCAAGAGAGATCATTTTGGAGCTTTAAGATTTGACTGCC 426  
QY 1274 CCATCGGATTTTGGACTCTCATGGGCCCTAGCCCTTTGTTTGGCCAAATTTTTCAT 1333  
DB 425 CCATCGGATTTTGGACTCTCATGGGCCCTAGCCCTTTGTTTGGCCAAATTTTTCAT 366  
QY 1334 TTGGAATGCCCTATTTATCCCAATGCCCTGACTCATTCATTTGAT 1376  
DB 365 TTGGAATGCCCTATTTATCCCAATGCCCTGACTCATTCATTTGAT 323

RESULT 6  
AAS03689/c  
ID AAS03689 standard; DNA; 9236 BP.  
XX  
AC AAS03689;  
XX  
DT 29-AUG-2001 (first entry)  
XX  
DE Rhesus gene locus: upstream Rhesus box of D-positives.  
XX  
KW Rhesus box; RHD positive; sequence length polymorphism; SSP; RHD; SMP1;  
KW RHCE; Rh negative; blood group typing; blood transfusion; antigen C;  
KW haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200132702-A2.  
XX  
PD 10-MAY-2001.  
XX  
PF 31-OCT-2000; 2000WO-EP10745.  
XX  
PR 02-NOV-1999; 99EP-0121686.  
XX  
PR 31-MAY-2000; 2000EP-0111696.  
XX  
PA (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTEMBERG.  
XX  
PI Flegel WA, Wagner FF;  
XX  
DR WPI; 2001-291052/30.  
XX  
PT New nucleic acid molecular structure, useful for detection of common  
PT RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1  
PT and RHCE genes -  
XX  
PS Disclosure; Fig 9; 135pp; English.  
XX

CC The sequence represents the coding sequence of Rhesus gene locus:  
CC upstream Rhesus box of D positives. The Rhesus genes locus  
CC comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the  
CC Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box  
CC and/or the downstream Rhesus box. The RHD and RHCE genes are located at  
CC chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the  
CC RHD deletion in the common RHD negative haplotypes. The sequence has  
CC been used to design primers which are useful for: (1) the specific  
CC detection of the common RHD positive haplotypes in D-negative  
CC individuals; (2) blood group typing; (3) determining whether a patient  
CC can be transfused with Rhd negative blood and whether blood is suitable  
CC for transfusion to patients who should not be exposed to antigen C; (4)  
CC assessing the risk of a Rhd negative mother of conceiving or carrying an  
CC Rhd positive foetus. Anti-D antibodies are useful for treating pregnant



QY 411 AACGATTCAGAGGTGACTTGGTGCTGTAAAGGCATTCAGTTTCATAAAGGAGGCAGA 470  
Db 5993 AACGATTCAGAGGTGACTTGGTGCTGTAAAGGCATTCAGTTTCATAAAGGAGGCAGA 5934  
QY 471 GCATAGAGTTCAGAAAAATTCACCCCTGACAAATGTGATAAAAAAGAAAA--CCCATTT 528  
Db 5933 GCATAGAGTTCAGAAAAATTCACCCCTGACAAATGTGATAAAAAAGAAAAATTCACATTT 5874  
QY 529 TCTGAGGGAAATTCAGGTGCTGCAGAAATTTGCATATGTAATGAGGAGCTGAATGTT 588  
Db 5873 TCTGAGGAGAAATTCAGGTGCTGCAGAAATTTGCATGAGTAA--CAGGAGCCAAATGCT 5815  
QY 589 AATCTCAGACAAATGGGAAATATCTCCTGGACATGTACAGGTCTTCACAGCAGTCC 648  
Db 5814 AATCTCAGACAAATGGGAAATATCTCAGGGATTCAGAGGTCTTCATGCGAACCC 5755  
QY 649 ATCAATCACTGCGCTGGAGGCT--AGGAGAAATGGTTTTCTGGGACAGGCCAGGGTC 707  
Db 5754 CTCCCATCAGGTCCAGAGGTATCAGGAAATATGGTTTTCTGGCAGGCCCGGGGTC 5695  
QY 708 CCGTGCTGTGCGACCTAGAGACTTGTGCCCTGTGTCCAGTTA----- 754  
Db 5694 CTCATGCTGTGCGACCTAGGACTTGTGCCCTGTGCATCCCGACCACTCCCAACCATGA 5635  
QY 755 -----ATTGAGTGTGCTTTTCAGAGGGTGCAGCCCAAGCCCTT 793  
Db 5634 CTGACGGGAGCAAGTAGACTTGGGCTGTAGCTCGGGGAGTGCAGCCCAAGCCCTT 5575  
QY 794 GCGAGCTTCCAGTGGTGTGAGCTGTGGGTGCAAGAAATGCAAGATTCAGATTTGAGTTGG 853  
Db 5574 GACAGCTTCCATGGTGTGAGATGCGAGTGCAGAAATGCAAGATTTGGGGTTTGA 5515  
QY 854 AACCTCCAAATCAGATTTCAAGAGATATATGAAACCCCTGGATGCCAGGCAGAAATTTG 913  
Db 5514 AACCTTCGCTAGATTAAGAGAGATGTGCGGAAATGCTGGATGCCAGTCAAGATTTG 5455  
QY 914 CTGTAGGGTGGGGTCTCATCGAGAACTCTGCAAGGGTAGTACAAAGAGGAAATGTTG 973  
Db 5454 CTGACGAGCAGGGCCCTCATGGAGATCTCTGCCAGGCGAGTGCAGAAAGGAAATGTTG 5395  
QY 974 GTGGGAGCCCCACACAGAGTCCCGAGTGGGCTCCATCTAGTAGAGCTGTGAGAGAA 1033  
Db 5394 GTTCAGAGACCCACACAGTCCCTACTGGGGCCACCTAGTGGAGCTGTGAGAGAG 5335  
QY 1034 GTCCACCATCTCCAGACTCCAGAGGTGAGATCCACTGACAGCTTGACAGCTGTGCCCTG 1093  
Db 5334 G-----TCCTCCAGACCCAGAAATGATAGATCCACGACAGCTTGACCCGTGACCTG 5282  
QY 1094 AAAATCCAGACACTCAGTCCAGCCCTGTGAAAGCAGCAGGGATGGAGTCTGTACCTT 1153  
Db 5281 GAAAAGCTCAGACACTCAATGCCAGCCCATGAAAGCAGCTGAGAGGGAGGCTGTACCTT 5222  
QY 1154 ACAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTCTTGCAATTTGATG 1213  
Db 5221 GCAAGGTACAGGGCAGAGCTGCCCAAGACCATGGGAACCCACCTTCCATCTGCGTG 5162  
QY 1214 ACCTGGACGTGAGACATGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTTGAGTGGC 1273  
Db 5161 ACCTGGATGTGAGATGTGGAGTCAAGAGAGATCATTTTGGAGCTTTAAGATTTGAGTGGC 5102  
QY 1274 CCATGGATTTCCGACTTATATATGGGGCCCGTACCCCTTTGTTTGGCCAAATTTTTCAT 1333  
Db 5101 CCATGGATTTTGGACTCTCATGGGCGCTGTAGCCCTCTTGTGTTTGTACCAATTTATCCCAT 5042  
QY 1334 TTGGGACTGCCCTATTTACCAATGCCCTGACTCCATGTTAT 1376  
Db 5041 TTGGAATGGCTGTATTTACCAATGCCCTGTACCCCATTTGTTAT 4999

RESULT 8  
AAS03688/c  
ID AAS03688 standard; DNA; 9241 BP.  
XX

AC AAS03688;  
DT 29-AUG-2001 (first entry)  
XX Rhesus gene locus: hybrid Rhesus box of RHD negatives.  
DE Rhesus box: RHD positive; sequence length polymorphism; SSP; RHD; SMP1;  
KW RHCE; Rh negative; blood group typing; blood transfusion; antigen C;  
KW haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.  
OS Homo sapiens.  
XX WO200132702-A2.  
XX 10-MAY-2001.  
XX 31-OCT-2000; 2000WO-EP10745.  
XX 02-NOV-1999; 99EP-0121686.  
PR 31-MAY-2000; 2000EP-0111696.  
XX (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTEMBERG.  
XX Flegel WA, Wagner FF;  
PI WPI; 2001-291052/30.  
DR New nucleic acid molecular structure, useful for detection of common  
PT RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1  
PT and RHCE genes -  
XX Disclosure: Fig 8; 135pp; English.  
XX The sequence represents the coding sequence of Rhesus gene locus:  
CC hybrid Rhesus box of RHD negatives. The Rhesus genes locus  
CC comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the  
CC Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box  
CC and/or the downstream Rhesus box. The RHD and RHCE genes are located at  
CC chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the  
CC RHD deletion in the common RHD negative haplotypes. The sequence has  
CC been used to design primers which are useful for: (1) the specific  
CC detection of the common RHD positive haplotypes in D-negative  
CC individuals; (2) blood group typing; (3) determining whether a patient  
CC can be transfused with RHD negative blood and whether blood is suitable  
CC for transfusion to patients who should not be exposed to antigen C; (4)  
CC assessing the risk of a Rhd negative mother of conceiving or carrying an  
CC Rhd positive foetus. Anti-D antibodies are useful for treating pregnant  
CC women who are Rhesus D negative, where the foetus is not homozygous for  
CC the RHD gene to treat or prevent haemolytic disease of the newborn.  
XX  
SQ Sequence 9241 BP; 2454 A; 2320 C; 2014 G; 2453 T; 0 other;  
Query Match 46.5%; Score 639.8; DB 22; Length 9241;  
Best Local Similarity 79.1%; Pred. No. 3e-189;  
Matches 841; Conservative 0; Mismatches 177; Indels 45; Gaps 5;  
QY 351 GAATATTTGGCTAATCAGGAAGCAAAATTTTGGCTAATAAGGAAGAAATTTCTAAGCAGCA 410  
Db 6053 GAACITTTGAACITTCAGAGAGATGATTTAGAGTATCTGGCAGGAGAAATTTCTAAGCAGCA 5994  
QY 411 AAGCATTCAGAGGTGACTTGGTGCTGTGTTAAAGGCATTCAGTTTCATAGGAGGCAGCA 470  
Db 5993 AAGCATTCAGAGGTGACTTGGTGCTGTGTTAAAGGCATTCAGTTTCATAGGAGGCAGCA 5934  
QY 471 GCATAGAGTTCAGAAAAATTCACCCCTGACAAATGTGATAAAAAAGAAAA--CCCATTT 528  
Db 5933 GCATAGAGTTCAGAAAAATTCACCCCTGACAAATGTGATAAAAAAGAAAAATTCACATTT 5874  
QY 529 TCTGAGGGAAATTCAGGTGCTGCAGAAATTTGCATATGTAATGAGGAGCTGAATGTT 588  
Db 5873 TCTGAGGAGAAATTCAGGTGCTGCAGAAATTTGCATGAGTAA--CAGGAGCCAAATGCT 5815  
QY 589 AATCTCAGACAAATGGGAAATATCTCCTGGACATGTACAGGTCTTCACAGCAGTCC 648









```
FT FT /*tag= b
FT FT /number= 3
FT FT /cons_splice= (5'site:YES,3'site:NO)
FT FT replace(64331,G)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT 166228..166322
FT FT /*tag= b
FT FT /number= 4
FT FT 166323..168001
FT FT /*tag= b
FT FT /number= 4
FT FT replace(167950,G)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(167989,G)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT 168002..168120
FT FT /*tag= b
FT FT /number= 5
FT FT replace(168054,G)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT 168121..169542
FT FT /*tag= b
FT FT /number= 5
FT FT /cons_splice= (5'site:YES,3'site:NO)
FT FT 169543..169825
FT FT /*tag= b
FT FT /number= 6
FT FT replace(169812,G)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(169823,G)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT 169826..169866
FT FT /*tag= b
FT FT /number= 6
FT FT /cons_splice= (5'site:YES,3'site:NO)
FT FT 169867..170678
FT FT /*tag= b
FT FT /number= 7
FT FT replace(170035,A)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(170068,T)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(170256,C)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(170368,G)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(170487,C)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT 170679..204911
FT FT /*tag= b
FT FT /number= 7
FT FT 204912..205102
FT FT /*tag= b
FT FT /number= 8
FT FT 205103..242969
FT FT /*tag= b
FT FT /number= 8
FT FT 242970..243086
FT FT /*tag= b
FT FT /number= 9
FT FT replace(243055,T)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT 243087..306167
FT FT /*tag= b
FT FT /number= 9
FT FT replace(243187,C)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT 306168..306503
FT FT /*tag= b
FT FT /number= 10
FT FT replace(306292,A)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(306382,G)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT 306504..373639
FT FT /*tag= b
FT FT /number= 10
FT FT 373640..373778
FT FT /*tag= b
FT FT /number= 11
FT FT 373779..422963
FT FT /*tag= b
FT FT /number= 11
FT FT 422964..423097
FT FT /*tag= b
FT FT /number= 12
FT FT replace(423067,C)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT 423098..456353
FT FT /*tag= b
FT FT /number= 12
FT FT replace(423149,G)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(423163,G)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(423220,A)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(423232,G)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(423258,G)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT 456354..456537
FT FT /*tag= b
FT FT /number= 13
FT FT 456538..460700
FT FT /*tag= b
FT FT /number= 13
FT FT replace(459706,C)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(459832,A)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(459913,G)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(460024,G)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT 460056..460056,T
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(460159,G)
FT FT /*tag= a
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT
```

```
FT variation replace(460553,C)
FT /*tag= a
FT /*standard_name= "Single nucleotide polymorphism"
FT variation replace(460564,A)
FT /*tag= a
FT /*standard_name= "Single nucleotide polymorphism"

Query Match 44.9%; Score 618.2; DB 24; Length 465237;
Best Local Similarity 78.2%; Pred. No. 1.9e-181;
Matches 847; Conservative 0; Mismatches 193; Indels 43; Gaps 7;

QY 334 CTCGATCATGCAAGGTGAATATTGGCTAATCAGAGCAAAATTTTGGCTAATAAGGAA 393
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 438730 CTCGCTAGAGATTGTGGAAATTGAACTTGAGAGAAATGATTTGGGGTACCTGGTAAAA 438789
QY 394 GAAATTCTTAAGCAGCAAAAGCATTTCAAGAGGTGACCTTGGGTGCTGTTAAAGGCATTCAGT 453
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 438790 GAAATTCTTAAGCAGCAAAACATTTCAAAAGGTGACCTTGGGTGTTGTAAAGCATTCGTG 438849
QY 454 TTCATTAAGGAGCAGCAGCATTAAGATTCAGAAAAATTTGCACCCTGCACAAATGTGATAAAA 513
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 438850 TT--TAAAGAGAAACAGCATAAAAGTTTCAGAAAAATTTGCAGCCTGATGATGCAGTAGGA 438907
QY 514 AAGAAAAACCCA--TTTCTCGAGGGGAAATTCAGCTGGCTGCAGAAATTTGCATATGTAA 572
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 438908 AAGAAAAACCAATTTTTCAGGAGAAATTCAGCTGGCTGCAGAAATTTGCATAAGTAA 438967
QY 573 TGAGGAGCTGAATGTTAATCCTCAAGACAAATGGGAAAAATATCTCCTGGACATGTGCAGAG 632
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 438968 CAAGGAGCCAAATGTTAATCCCAAGACAAATGGGAAAAATGTCCTCAGAGCATGTCTAG 439027
QY 633 GTCTTTCACAGCACTCAATCAATCACTGGCTGGAGGCCTAGGAG---AAAATGGTTTT 688
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 439028 GTCTTTCATGCGACCCCTCCCATCACAGACCCCGAGCCTAGGAGGAAAAAACAGTTTT 439087
QY 689 GTGCGACAGCCAGAGTCCCTGCTGCTGTGCAGCTAGAGACTTGGTGCCCTGTGTC 748
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 439088 GTGGGCGAGTCCAGGGTCCCAATGCTGTGTGCAGCCTAGGAACCTTGGTGCCCTGCATCT 439147
QY 749 CAGTTAATTCAGCTGT-----GGCTTCAGAGGGGTGCAAGCC 784
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 439148 CAGCTGCTCAGCTATTGCTAAAGGGGTGAGGTACCACGGTTTCAGAGGTGCAAGCC 439207
QY 785 CCAAGCCTTGGCAGCTTCCAAGTGGTGTGAGCCTGTGGGTGCAAAAGAGTCAAGAAATTG 844
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 439208 CCAAGCCTTGGCAGCTTTCATGTGGTGTGAGCCTGTGTGTACACAGAAAGTTAAGAAATTG 439267
QY 845 AGCTTTGGGAACCTCCAATCAGATTTTCAGAAAGATATGGAACCCCTGGATGCCAGGC 904
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 439268 AGGTTTGGGAACCTCCACTTATATTTCAGAAAGATAGTGGAAATGCTGGATACCCAGGC 439327
QY 905 AGAAGTTTGTGTAGGGTGGGTGCTCATGGAGAACCTCTGCAAGGGTAGTACAAAAAGG, 964
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 439328 AAACATTTGCTGAGAGGTGGGGCCCTCATGGAGGGCCCTGCTAGGCCAATCAGGAAGG 439387
QY 965 GAAATGTGGGTGGGAGCCCCACACAGAGTCCCAAGTGGGGTCCATCTAGTAGAGCTG 1024
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 439388 GAAATGTGGGTGGGAGCCCCACACAGAGTCCCAAGTGGGGTCCCTAGTGGAGCTG 439447
QY 1025 TGAGNAGAGTCCACCATCTCCAGACTCCAGAGGAGGTAGATCCACTGCAGCTTGCAGC 1084
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 439448 TAAGAGGAGGACCACTGTCTCTCAGACCCGAGAAATAGTAGATCCACTGCAGCTTGCACC 439507
QY 1085 ATGTGCTGAAAAATCCACAGACACTCAGTGCCAGGCTGTGAAAGCA-----GCA 1134
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 439508 ATGTGCTGAAAGGCCACAGACACTCAAGCCAGCCTGTGAAAGCAGTCAGGGTTGGAG 439567
QY 1135 GGGATGGAGTCTGTACCTACAAAAACCGTAGTGGCAGAGCTGACCAAGACCCGTGGGAATC 1194
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 439568 GTGGTGGTGTATACCTTATAAAGCCACAGAGGGCAGAGCTGCCCCAAGACTATGGGAACC 439627
QY 1195 TACCTCTTGCATGTCATGACCTGGAGCTGAGACATGGAGTCAAAAGAGATCATTTTGGGA 1254
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

QY	1314	TTTTGGCAATTTTCCATTGGAAGCTGCGATTATTACCAATGCTGTACTGTCATTG	1373
Db	439747	TTTTGGCAATTTTCCATTGGAAGCTGCGATTATTACCAATGCTTAACCCGCAATTG	439806
QY	1374	TAT 1376	
Db	439807	TAT 439809	
RESULT 13			
	AAD05134		
ID	AAD05134	standard; cDNA; 3203 BP.	
XX	AAD05134;		
AC	AAD05134;		
XX			
DT	17-JUL-2001	(first entry)	
XX			
DE	Human secreted protein-encoding gene 14 CDNA clone HISBF60, SEQ ID NO:24		
XX			
XX	Human; secreted protein; proliferative disorder; cancer; tumour;		
KW	fetal abnormality; developmental abnormality; haematopoietic disorder;		
KW	immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;		
KW	inflammation; allergy; neurological disorder; Alzheimer's disease;		
KW	Parkinson's disease; cognitive disorder; schizophrenia; asthma;		
KW	skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;		
KW	cardiovascular disorder; angioleptic disorder; kidney disorder;		
KW	gastrointestinal disorder; pregnancy-related disorder;		
KW	endocrine disorder; infection; wound healing; vulnery;		
KW	cell culture; chemotaxis; food additive; gene therapy;		
KW	binding partner identification; ss.		
XX			
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	153..356	
FT		/*tag= a	
FT		/product= "Human secreted protein"	
FT	sig_peptide	153..215	
FT		/*tag= b	
FT	mat_peptide	216..353	
FT		/*tag= c	
FT		/product= "Human mature secreted protein"	
XX			
PN	WO200134769-A2.		
XX			
PD	17-MAY-2001.		
XX			
PF	01-NOV-2000; 2000WO-US30040.		
XX			
PR	05-NOV-1999; 99US-0163580.		
PR	30-JUN-2000; 2000US-0215130.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Ruben SM, Komatsoulis GA, Wei P, Baker KP, Fiscella M;		
XX			
DR	WPI; 2001-308781/32.		
DR	P-PSDB; AAE01245.		
XX			
PT	New isolated nucleic acid molecule encoding a human secreted protein is		
PT	used in preventing, treating or ameliorating a medical condition -		
XX			
PS	Claim 1; Page 408-409; 519pp; English.		
XX			
CC	AAD05121-AAD05203 represent cDNAs corresponding to 24 human secreted		
CC	protein genes, and AAE01232-AAE01311 represent the proteins they encode.		
CC	AAE01312-AAE01340 represent human secreted protein variants or fragments		
CC	The secreted proteins and their genes are useful for preventing,		
CC	treating or ameliorating medical conditions, e.g., by protein or gene		
CC	therapy. Pathological conditions can be diagnosed by determining the		
CC	amount of the new protein in a sample or by determining the presence of		
CC	mutations in the new genes. Specific uses are described for each of the		

24 genes, based on the tissues in which they are most highly expressed, and include developing products for the diagnosis or treatment of proliferative disorders, cancer, tumours, foetal and developmental abnormalities, haematopoietic disorders, diseases of the immune system, AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation, allergies, neurological disorders (e.g., Alzheimer's disease, Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders, oncogenic disorders, kidney disorders, gastrointestinal disorders, pregnancy-related disorders, endocrine disorders, and infections. The proteins can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues, to identify their cognate ligands or binding partners, and in chemotaxis, and can be used as a food additive or preservative to modify storage properties. Antibodies specific for a protein of the invention can be used in alleviating symptoms associated with the disorders mentioned above, and in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The present sequence represents a human secreted protein-encoding cDNA of the invention.

Sequence 3203 BP: 868 A; 650 C; 740 G; 943 T; 2 other;

Db	2256	TTTGGAGCCCCACACAGAAATCCTTACTGTGGGCAACCACTAGTGGAGCTGTGCGAAGAAGG	2315
Qy	1036	CCACCATCTCTCAGACTCCAGAAAGGCTAGATCCACTGACAGCTTGGCAGCATGTGCCCTGAA	1095
Db	2316	CCACCGTCTTCAGACCCAGAAATGCTAGATCCACGACAGCTTGTACCGTGCACCTGGA	2375
Qy	1096	AAATCCACAGACACTCAGTGCAGCCAGCTGTGAAGCAG-CAGGGATGGAGTCTGTACCCCTA	1154
Db	2376	AAAGCCACAGACCCATAACGCCAGCCCGTGAAGCAGCAGGAGTGGGACTATACCCCTG	2435
Qy	1155	AAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAACTACCTCTTGCATTTGTCTATGA	1214
Db	2436	TGAAGCCACAGGGCGAGAGCTGCCCAAGACTAAGGGAACTACCTCTTGCATCATTTGTA	2495
Qy	1215	CCTGCAGCTGAGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTTGACTGCCC	1274
Db	2496	CTGTGATGTGAGACATGGAGTCAAGAGGATCATTTTGAACGATTAATTTGACTGCCT	2555
Qy	1275	CAGTGGATTCGGACTTATATGGGGCCCGT-ACCCCTTTGTTTGGCCAATTTTTTCCAT	1333
Db	2556	CGCTGGATTTTGGACTTGCATGGGGCTTGTAAAGCCCTTTGTTGGCAGTTTCTCCCAT	2615
Qy	1334	TTGGAATGCCGTATTTACCCCAATGCCCTGCTACCTCCATTGTATG	1377
Db	2616	TTGGAATGSGCTGATTTTACCCCATTACCTGTATCCCCATTGTTATG	2659
RESULT 14			
AAL03634			
ID	AAL03634 standard; DNA; 6063 BP.		
XX	AAL03634;		
XX			
DT	21-NOV-2001 (first entry)		
XX	Human reproductive system related antigen DNA SEQ ID NO: 6322.		
XX	Human; reproductive system related antigen; reproductive system disorder;		
KW	cancer; gene therapy; ds.		
XX	Homo sapiens.		
OS	Homo sapiens.		
PN	WO200155320-A2.		
XX	02-AUG-2001.		
XX			
PF	17-JAN-2001; 2001WO-US01339.		
XX			
PR	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	24-FEB-2000; 2000US-0184664.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0198974.		
PR	17-MAR-2000; 2000US-0190076.		
PR	18-MAR-2000; 2000US-0198123.		
PR	19-MAY-2000; 2000US-0205515.		
PR	07-JUN-2000; 2000US-0209467.		
PR	28-JUN-2000; 2000US-0214886.		
PR	30-JUN-2000; 2000US-0215135.		
PR	07-JUL-2000; 2000US-0216647.		
PR	07-JUL-2000; 2000US-0216880.		
PR	11-JUL-2000; 2000US-0217487.		
PR	11-JUL-2000; 2000US-0217496.		
PR	14-JUL-2000; 2000US-0218290.		
PR	26-JUL-2000; 2000US-0220963.		
PR	26-JUL-2000; 2000US-0220964.		
PR	14-AUG-2000; 2000US-0224518.		
PR	14-AUG-2000; 2000US-0224519.		
PR	14-AUG-2000; 2000US-0225213.		
PR	14-AUG-2000; 2000US-0225214.		
PR	14-AUG-2000; 2000US-0225266.		
PR	14-AUG-2000; 2000US-0225267.		



Db 3559 GCATATAAAATTTGGAAAAATTTGCAGCCAGTTGATGCAGCAGAGAGAGAAACCTGTTTTT 3618  
QY 531 TGAGGGGAAATCAAGCTGGCTGCAGAAATTTGCATATGTAATAGGAGCTGAATGTTAA 590  
Db 3619 TGAGGAGAAATCAAGCTGGCTGCAGAAATTTGCATAAGTTACAGGGAGCTGAATGTTAA 3678  
QY 591 TCGTCAAGACAATGGGGAAAAATATCTCTCGGAGATGTCTCAGAGGCTCTTATGTCAGCCCT 650  
Db 3679 TCTCCAAGACAATGGGGAAAAATGTTCTCCAGGCGATGTCACAGGTCTTATGTCAGCCCT 3738  
QY 651 CAAATCACTGGCCCTGGAGCCCTAGAG---AAATGGTTTTTGGGACAGGCCAGGGTC 707  
Db 3739 CCATCAGAGCCCTGAAGACTAGGAGGAAAAATGGTTTTTGGGCCAGGCCAGGGTC 3798  
QY 708 CTGTGCTGTGTGCAGCCCTAGAGACTTGTGTCCTGTGCTCCAGTTAATTACG----- 760  
Db 3799 CCCATGCTGTGTGCAGCCCTAGGAGCTTGTGTCCTGTGCTCCAGCTCCAGCTGCTCCAGCCATTGC 3858  
QY 761 -----CTGTGGCTTCAGAGGGTGCAGCCCAAGCCCAAGCCTTGG 795  
Db 3859 TAAAGGGCGAGGTACAGCTCTGCCCGTGTTCAGAGGGTGCAGCCCAAGCCTTGG 3918  
QY 796 CAGCTTCCAAGTGGTGTTCAGCCTGTGGGTGCAAGAAGTCAAGATTCAGGTTGGGAA 855  
Db 3919 CAGCTTCCAAGTGGTGTTCAGCCTGTGGGTGCAAGAAGTCAAGATTCAGGTTGGGAA 3978  
QY 856 CTCCAATCAGATTTTCAGAGAATATATGAAACCCCTGGATGCCAGGCAGAGTTTGGT 915  
Db 3979 CCTCCATCTAGATTTTCAGAGAATATATGAAACCCCTGGATGCCAGGCAGAGTTTGGT 4038  
QY 916 GTAGGGGTGGGTCTCTATGAGAGACCTCTGCAAGGGTAGTACAAAAGGAAATGTTGG 975  
Db 4039 CAGGGGCGAGAGCCTTCATGAGAGACCTCTGCTAGGGCAGT-GTGAAGGAAATGTTGGG 4097  
QY 976 TGGAGCCCCACACAGAGTCCCCAGTGGGCTCCATCTAGTAGAGCTGTGAGAGAACT 1035  
Db 4098 TGGAGCCCCACACAGAGATCCCTACTTGGGGCACCACCTAGTGGAGCTGTGCAAGAAG 4157  
QY 1036 CCACATCTCCAGACTCCAGAGGTTAGATCCACTGCAGCTTGCAGCATGTGCTGAA 1095  
Db 4158 CCACGCTCTTCAGACCCCAAGATGTTAGATCCAGCAGAGCTTGTACCGTCACCTGGA 4217  
QY 1096 AAATCCACAGACACTAGTGCAGCCTGTGAAAGCAG-CAGGGATGGAGTCTGTACCCCTA 1154  
Db 4218 AAAGCCACAGACCTAAAGCCAGCCGCTGAAAGCAGCAGGAGTGGGAGCTATACCCCTG 4277  
QY 1155 CAAACCGTAGTGGAGAGCTGACCAAGACCGTGGGAATCTACCTTGCATGTGCATGA 1214  
Db 4278 TGAAGCCAGGGGCGAGAGCTGCCAGACTAAGGGAACCTACCTCTTGCATCATTTGA 4337  
QY 1215 CTGGACGCTGAGACATGGAGTCAAAAGAGATCATTTTGGAGGCTTTAAGATTTGACTGCC 1274  
Db 4338 CTGGATGTGAGACATGGAGTCAGAGGAGATCATTTTGGAACTTATATATTGACTGCTC 4397  
QY 1275 CACTGGATTTGGAGCTTATATGGGCCCT-ACCCGTTTGGTTTGGCCAAATTTTTCAT 1333  
Db 4398 CCGTGGATTTGGAGCTGATGATGGGCTGTGAACGCTTGTGTTGGGCACTTCTCCCAT 4457  
QY 1334 TTGGAACCTCGCTATTATCCCAATGCTGTACCTCATTTGATG 1377  
Db 4458 TTGGAATGCTGTATTATCCCAATGCTGTATCCCATTTACCTGTATCCCATTTGATG 4501

RESULT 15

AAH13678  
ID AAH13678 standard; cdna; 1736 BP.

XX AAH13678;

XX 26-JUN-2001 (first entry)

DE Human cdna sequence SEQ ID NO:10539.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

PN EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 93JP-0248036.

PR 27-AUG-1999; 95JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

full-length cDNAs defined in the specification, and for the detection

and/or diagnosis of the abnormality of the proteins encoded by the

full-length cDNAs.

PS Claim 8; SEQ ID 10539; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

full-length cDNAs defined in the specification. Where a primer set

comprises: (a) an oligo-dT primer and an oligonucleotide complementary

to the complementary strand of a polynucleotide which comprises one of

the 5602 nucleotide sequences defined in the specification, where the

oligonucleotide comprises at least 15 nucleotides; or (b) a combination

of an oligonucleotide comprising a sequence complementary to the

complementary strand of a polynucleotide which comprises a 5'-end

sequence and an oligonucleotide comprising a sequence complementary to a

polynucleotide which comprises a 3'-end sequence, where the

oligonucleotide comprises at least 15 nucleotides and the combination of

the 5'-end sequence/3'-end sequence is selected from those defined in

the specification. The primer sets can be used in antisense therapy and

in gene therapy. The primers are useful for synthesizing polynucleotides,

particularly full-length cDNAs. The primers are also useful for the

detection and/or diagnosis of the abnormality of the proteins encoded by

the full-length cDNAs. The primers allow obtaining of the full-length

cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to

AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632

represent oligonucleotides, all of which are used in the exemplification

of the present invention.

QY Sequence 1736 BP; 449 A; 372 C; 453 G; 462 T; 0 other;

Query Match 43.8%; Score 602.8; DB 22; Length 1736;

Best Local Similarity 79.1%; Pred. No. 4.6e-178;

Matches 810; Conservative 0; Mismatches 167; Indels 47; Gaps 6;

QY 391 GAAGAAATTTCTAAGCAGCAAGCATTCAAGAGGTGCTGGTGTCTGTAAAGGCATTC 450

Db 111 GAAGAAATTTCTAAGCAGCAAGCATTCAAGAGGTGCTGGTGTCTGTAAAGGCATTT 170

QY 451 AGTTTCATAAGGAGGAGCAGCATAGAGTTCAGAAAAATTTGCACCATGCAATGTGATA 510

Db 171 AGTTTATAGGAGGAGCAGCATAGAGTTCAGAAAAATTTGCAGCATTTACTATGCGATA 230

QY 511 AAAAAAGAAAAACCCATTTTCTCAGGGGAAATTCAGCTGGCTGCAGAAATTTGTCATATGT 570

Db 231 GACAAGAAAAACCCATTTTCTCAGGGAGAGATTCAAGCCAGCTGCGGAAATTTGTGTAAGT 290

QY 571 AATGAGGAGCTGAATGTTAATCTCTCAAGACAATGGGAAAAATATCTCTCCCTGGAGCATGTCAG 630

Db 291 AGCAAGGAGCCTAATGTTAGTCCCAAGACCATGGGGAAGATGCTCCAGACCATGTCTCAG 350  
Qy 631 AGGTCTTCACAGCAGTCCATCAAAATCACTGGCCTGGAGCCCTAGGAGAAAATGGTTTGT 690  
Db 351 AGACCTTCACCACAGCCCTCCTATCACAGGCCCAAGAGGAAAAAGTGGTTTGT 405  
Qy 691 GGGACAGGCCAGGCTCCTGTGCTGTGTGAGCCTAGAGACTTGGTCCCTGTGTCCCA 750  
Db 406 GGGCTGTGCTCCAGGCTCCCTTGTGTGTGCAACCTAGGATGTGGCACCCCTGTGTCCCA 465  
Qy 751 GTTAATTC-----AGCTGTGCTTCACAG 774  
Db 466 GCTGCTGTTTCAGCTGTGGCTGAAAGGGGCCAATGTACAGTCAAGTTGTGGCCTCAGAG 525  
Qy 775 GGTGCAAGCCCCAAGCCTTGGCAGCTTCCAAGTGGTGTGAGCCTGTGGGTGCAAAAGAG 834  
Db 526 GGTGAAGCCCCAAGCCTTGGCAGGCTCCACATGGTGTGAGCCTGACAGTGCACAGAAG 585  
Qy 835 TC-AAGAATTGAGGTTGGGAACCTCCAATCAGATTTCAGAAGATATATGGAACCCCTG 893  
Db 586 TCAAGAATTGAGGTTGGGAACCTCTACCTAGATTTTC---AGATGTATGGAATGTCTA 642  
Qy 894 GATGCCAGGAGGAGTGTGTGAGGGTGGGTCTCATGGAGAACCTCTGCAAGGCT 953  
Db 643 GATGCCAGGAGGAGTGTGTGAGGGTGGGTCTCATGGAGAACCTCTGCTAGGGC 701  
Qy 954 AGTACAAAAGGGAATGTTGGTGGGAGCCCCCACACAGAGTCCCCAGTGGGGCTCCATC 1013  
Db 702 AGTGTGAAGGGAATGTGGGTTGGAGCCCCCACACAGAGTCCCTACTGGGGCCTGCC 761  
Qy 1014 TAGTAGAGCTGTGAGAAGAAGTCCACCATCTCCAGACTCCAGAGGCTAGATCCACTGA 1073  
Db 762 TAGTAGAGCTGTGAGAAGAGGCCATTGTCTTCTAGACCCCAAGATGGCAGACCCACTGA 821  
Qy 1074 CAGCTTGACAGATGTGCTGAAATCCACACAGACTCAGTCCCAGCCTGTCAAGCAGC 1133  
Db 822 CAGCTTGACCATGTGCTGGAAAGCTGCAGACACTCAACGCCAGCCCGTGAAGCATC 881  
Qy 1134 AGGATGGAGTCTGTACCTACAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAAT 1193  
Db 882 CGGGAGGAGGCTGTACCTGCCAAGCCACAGGGCAGAGCTACCCAAGACCATGGGAAC 941  
Qy 1194 CTACCTCTTGCAATGTCTATGACCTGGAGCTGAGACATGGAGTCAAAAGAGATCATTTTG 1253  
Db 942 TCACCTCTTGCAATGAGTGTGACCTGGATTGAGACCTGGAATCAAAAGGAGATCATTTTG 1001  
Qy 1254 AGCTTTAGATTTGACTGCCCTGAGTTTCGGACCTTATATGGGCCCGT-ACCCCTTT 1312  
Db 1002 AGCTTTAACAATTTGACTGCCCTGAAATCTGGACTTGCATGGGACCCCTAACCCCAT 1061  
Qy 1313 GTTTTGGCAATTTTTCATTGGAACTGCCGTATTTACCAATGCCCTGTACCTCCATT 1372  
Db 1062 GTTTTGGCAATTTCTCTCATTTGGAATGGCTGTATTTACCCAATACCTGCACACCCATT 1121  
Qy 1373 GTAT 1376  
Db 1122 GTAT 1125

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:16:26 ; Search time 4390.26 Seconds  
(without alignments)  
9128.055 Million cell updates/sec

Title: US-09-997-610-1-copy\_2\_1378

Perfect score: 1377

Sequence: 1 atagtggtcatacctgtctt.....gctgtacctccattgtatg 1377

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pi.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pi.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rnd.\*

36: em\_htg\_mam.\*

37: em\_htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1281	93.0	145880	9	HS302D9	Z82198 Human DNA s
C 2	695.2	50.5	193092	9	AC017063	AC017063 Homo sapi
C 3	691	50.2	124518	9	AL138962	AL138962 Human DNA
C 4	690.8	50.2	77405	2	AL353634	AL353634 Homo sapi
5	690.2	50.1	40714	9	HSU212C1	Z69722 Human DNA s
6	689	50.0	123098	9	AC008799	AC008799 Homo sapi
7	689	50.0	154090	9	AC025577	AC025577 Homo sapi
8	689	50.0	174874	9	AC123595	AC123595 Homo sapi
9	689	50.0	186660	2	AC026107	AC026107 Homo sapi
C 10	689	50.0	189768	2	AC044889	AC044889 Homo sapi
C 11	684.2	49.7	170368	9	AC091982	AC091982 Homo sapi
12	683.6	49.6	67984	9	AC010312	AC010312 Homo sapi
C 13	683.2	49.6	168502	9	AC091005	AC091005 Homo sapi
14	681.6	49.5	131215	9	AC079614	AC079614 Homo sapi
C 15	680	49.4	73390	9	AC012038	AC012038 Homo sapi
C 16	679.8	49.4	152544	9	CNS05TEJ	AL359232 Human chr
C 17	679.8	49.4	203726	2	AC011882	AC011882 Homo sapi
18	679	49.3	173480	9	CNS00M8T	AL079343 Human chr
C 19	678.8	49.3	146743	2	AC093588	AC093588 Homo sapi
C 20	678.8	49.3	166679	9	AC079899	AC079899 Homo sapi
C 21	678.8	49.3	176426	9	AC007370	AC007370 Homo sapi
22	678.6	49.3	38235	9	AC004559	AC004559 Homo sapi
23	678	49.2	3417	9	AK091968	AK091968 Homo sapi
24	677.8	49.2	91323	9	AK106901	AK106901 Homo sapi
C 25	677.2	49.2	124531	9	HSJ633H17	AL049710 Human DNA
C 26	676.8	49.2	187898	2	AC016715	AC016715 Homo sapi
C 27	676.6	49.1	110000	2	AL732359_2	Continuation (3 of
C 28	676.6	49.1	138271	9	AC008360	AC008360 Homo sapi
29	676.6	49.1	147971	9	HS431P23	AL009178 Human DNA
C 30	676.4	49.1	77774	9	AP000339	AP000339 Homo sapi
31	676.4	49.1	100000	9	AP000217	AP000217 Homo sapi
32	676.4	49.1	340000	9	AP001760	AP001760 Homo sapi
33	675.8	49.1	107885	9	AC006389	AC006389 Homo sapi
34	675.8	49.1	150332	9	AC004921	AC004921 Homo sapi
C 35	675.6	49.1	181842	2	AL391823	AL391823 Homo sapi
C 36	675.4	49.0	94730	2	AP000230	AP000230 Homo sapi
37	675.4	49.0	100000	9	AP000144	AP000144 Homo sapi
38	675.4	49.0	100634	9	AP001594	AP001594 Homo sapi
39	675.4	49.0	340000	9	AP001695	AP001695 Homo sapi
C 40	675.2	49.0	155764	9	AC011238	AC011238 Homo sapi
C 41	675	49.0	182972	2	AC023550	AC023550 Homo sapi
C 42	674	48.9	174662	2	AC026036	AC026036 Homo sapi
C 43	673.8	48.9	123631	9	HS22F01	AL109967 Homo sapi
C 44	673.8	48.9	168863	9	AC011286	AC011286 Homo sapi
45	673.6	48.9	105692	9	AL451046	AL451046 Human DNA

ALIGNMENTS

RESULT 1

HS302D9

LOCUS

DEFINITION

Human DNA sequence from clone RPI-302D9 on chromosome 22 Contains

GSs, complete sequence.

282198

282198.2 GI:6572207

HTG.

HTG.

ORGANISM

Homo sapiens.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 145880)

AUTHORS

Bridgeman,A.

TITLE

Direct Submission

**JOURNAL**

Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk

**COMMENT**

On Dec 13, 1999 this sequence version replaced g1:3164067. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Emn:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>

RPI-302D9 is from the library RPCI-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR: PCYPAC2>

This sequence is the entire insert of clone RPI-302D9. The true left end of clone CFA-282F2 is at 69682 in this sequence. The true right end of clone CFA-415G2 is at 55167 in this sequence.

repeat_region	6647. .5685 /note="WADE1 repeat: matches 1. .23 of consensus"
repeat_region	6686. .6987 /note="AluX repeat: matches 1. .302 of consensus"
repeat_region	6988. .7036 /note="WADE1 repeat: matches 23. .77 of consensus"
repeat_region	7482. .7754 /note="AluJb repeat: matches 9. .290 of consensus"
repeat_region	7775. .8060 /note="AluJo repeat: matches 1. .295 of consensus"
repeat_region	8414. .8551 /note="L2 repeat: matches 2553. .2706 of consensus"
repeat_region	8914. .9030 /note="MIR repeat: matches 147. .262 of consensus"
repeat_region	9110. .9280 /note="MIR repeat: matches 91. .262 of consensus"
repeat_region	9283. .9412 /note="MIR repeat: matches 15. .144 of consensus"
repeat_region	9521. .9679 /note="FAM repeat: matches 3. .161 of consensus"
repeat_region	9820. .10225 /note="WSTB repeat: matches 2. .425 of consensus"
misc_feature	complement(10179. .10678) /note="match: GSS: Em:B56592"
misc_feature	complement(10204. .10728) /note="match: GSS: Em:AQ701486"
misc_feature	complement(10249. .10706) /note="match: GSS: Em:AQ225495"
repeat_region	10312. .10383 /note="MIR repeat: matches 79. .150 of consensus"
misc_feature	10718. .11310 /note="match: GSS: Em:B14024"
misc_feature	10784. .11201 /note="match: GSS: Em:B43656"
repeat_region	11838. .11946 /note="MIR repeat: matches 20. .137 of consensus"
repeat_region	12174. .12445 /note="L2 repeat: matches 1988. .2275 of consensus"
repeat_region	12444. .12642 /note="MIR repeat: matches 63. .241 of consensus"
misc_feature	13017. .13369 /note="match: STS: Em:G49301"
repeat_region	13331. .13397 /note="MIR repeat: matches 174. .244 of consensus"
repeat_region	13398. .13698 /note="AluSp repeat: matches 1. .302 of consensus"
repeat_region	13699. .13810 /note="MIR repeat: matches 76. .174 of consensus"
repeat_region	13806. .13919 /note="MIR repeat: matches 77. .189 of consensus"
repeat_region	13945. .14060 /note="MIR repeat: matches 24. .142 of consensus"
repeat_region	14061. .14367 /note="AluY repeat: matches 1. .301 of consensus"
repeat_region	14368. .14452 /note="MIR repeat: matches 141. .225 of consensus"
repeat_region	14589. .14679 /note="MIR repeat: matches 173. .262 of consensus"
misc_feature	14597. .15201 /note="match: GSS: Em:AQ553482"
misc_feature	14616. .15060 /note="match: GSS: Em:AQ370601"
repeat_region	14868. .15040 /note="MIR repeat: matches 49. .233 of consensus"
repeat_region	15071. .15188 /note="L2 repeat: matches 2112. .2239 of consensus"
repeat_region	15304. .15399 /note="MLTb repeat: matches 1. .99 of consensus"
repeat_region	15490. .15662 /note="AluSg1 repeat: matches 2. .114 of consensus"
repeat_region	15669. .15727 /note="MLTb repeat: matches 119. .178 of consensus"
repeat_region	15728. .15727 /note="MLTb repeat: matches 119. .178 of consensus"

```
repeat_region /note="AluSc repeat: matches 1. .299 of consensus"
16028. .16245
repeat_region /note="MLTIB repeat: matches 178. .390 of consensus"
16546. .16854
repeat_region /note="AluY repeat: matches 1. .300 of consensus"
18296. .18323
repeat_region /note="MSTA repeat: matches 2. .29 of consensus"
18324. .18392
repeat_region /note="MER66-internal repeat: matches 4919. .4993 of
consensus"
18393. .18712
repeat_region /note="AluYb repeat: matches 1. .311 of consensus"
18713. .19133
repeat_region /note="MER66-internal repeat: matches 4548. .4919 of
consensus"
complement(18872. .19230)
/misc_feature /note="match: GSS: Em:AO005063"
19251. .19719
/misc_feature /note="match: GSS: Em:BI4179"
19337. .20290
repeat_region /note="HERVH21 repeat: matches 4657. .5784 of consensus"
20317. .20382
repeat_region /note="33 copies 2 mer ta 68 conserved"
20513. .20666
repeat_region /note="77 copies 2 mer tt 70 conserved"
20682. .21008
repeat_region /note="AluSg1 repeat: matches 1. .306 of consensus"
21239. .21553
repeat_region /note="HUPRS-p3 repeat: matches 4410. .4713 of consensus"
21882. .22254
repeat_region /note="THS1B repeat: matches 1. .364 of consensus"
22302. .22537
repeat_region /note="MER66-internal repeat: matches 2186. .2417 of
consensus"
22538. .22850
repeat_region /note="AluSp repeat: matches 1. .313 of consensus"
22851. .23801
repeat_region /note="MER66-internal repeat: matches 1210. .2186 of
consensus"
23905. .23989
repeat_region /note="MER66-internal repeat: matches 3017. .3102 of
consensus"
Query Match 93.08; Score 1281; DB 9; Length 145880;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1284; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 89 TGGGGCTCCTGGTACCAGGTTTACCACAATATACAGGAGAATAAGTCAATGACAA 148
DB 37254 TGGAAATGCTCGTGGTCCCCAGGTTTACCACAATATACAGGAGAATAAGTCAATGACAA 37313

QY 149 AATGCCCTGTCTGATATAGAAAGGTCAGCCTTTACTGTGAAGCTCAGTGGAAACTTC 208
DB 37314 AATGCCCTGTCTGATATAGAAAGGTCAGCCTTTACTGTGAAGCTCAGTGGAAACTTC 37373

QY 209 CTCCTTCCTTCAAGCCCATCATCTTCACAGGGTCTGTGTACAATGCCAGAGGATTAA 268
DB 37374 CTCCTTCCTTCAAGCCCATCATCTTCACAGGGTCTGTGTACAATGCCAGAGGATTAA 37433

QY 269 AGGAGCCATGGGAGTCTTTCCTTGCAGGGTGCCTGGGAATTACTACTCAGCTTTGATG 328
DB 37434 AGGAGCCATGGGAGTCTTTCCTTGCAGGGTGCCTGGGAATTACTACTCAGCTTTGATG 37493

QY 329 TTGAGCTGCATCATTCGAAGTGTAATTTGGCTTAATGAGGAACAAATTTTGGCTAATA 388
DB 37494 TTGAGCTGCATCATTCGAAGTGTAATTTGGCTTAATGAGGAACAAATTTTGGCTAATA 37553

QY 389 AGGAAGAAATTTCTAAGCAGCAAAAGCATTCAGAGGTGACTTGGGTGCTGTGTAAAGGCAT 448
DB 37554 AGGAAGAAATTTCTAAGCAGCAAAAGCATTCAGAGGTGACTTGGGTGCTGTGTAAAGGCAT 37613

QY 449 TCAGTTTCATAGGAGGAGCAGAGCATAGAGTTTCAGAAAATTTGCACCCCTGACAAATGTGA 508
DB 37614 TCAGTTTCATAGGAGGAGCAGAGCATAGAGTTTCAGAAAATTTGCACCCCTGACAAATGTGA 37673
```

```
QY 509 TAAAAAGAAAAACCCATTTTCTGAGGGAAATTCAGCTGGCTGCAGAAATTTGCATAT 568
DB 37674 TAAAAAGAAAAACCCATTTTCTGAGGGAAATTCAGCTGGCTGCAGAAATTTGCATAT 37733
QY 569 GTAATGAGGAGCTGAATGTTAATCTCAAGCAATGGGAAATATCTCTCGACATGTC 628
DB 37734 GTAATGAGGAGCTGAATGTTAATCTCAAGCAATGGGAAATATCTCTCGACATGTC 37793
QY 629 AGAGGTCTTCAGCAGTCCATCAATCACCTGGCTGAGGGCTAGGAGAAATGTTT 688
DB 37794 AGAGGTCTTCAGCAGTCCATCAATCACCTGGCTGAGGGCTAGGAGAAATGTTT 37853
QY 689 GTGGACAGGCCCGGCTCCCTGTCTGTGTCAGCCCTAGAGACTTGTGTCCTGTGTC 748
DB 37854 GTGGACAGGCCCGGCTCCCTGTCTGTGTCAGCCCTAGAGACTTGTGTCCTGTGTC 37913
QY 749 CAGTTAATTCAGCTGTGGCTTCAGAGGGTGCAGAGCCCAAGCTTGGCAGCTTCAAGTG 808
DB 37914 CAGTTAATTCAGCTGTGGCTTCAGAGGGTGCAGAGCCCAAGCTTGGCAGCTTCAAGTG 37973
QY 809 GTGTTGAGCCTGTGGTGCAGAGGTCAAGAAATTCAGGTTTGGGAACCTCCAATCAGAT 868
DB 37974 GTGTTGAGCCTGTGGTGCAGAGGTCAAGAAATTCAGGTTTGGGAACCTCCAATCAGAT 38033
QY 869 TTCAGAAGATATATGAAACCCCTGGATGCCAGCAGAGAAATTTGCTGTAGGGTGGGT 928
DB 38034 TTCAGAAGATATATGAAACCCCTGGATGCCAGCAGAGAAATTTGCTGTAGGGTGGGT 38093
QY 929 CCTATGAGAGAACCTCTCCAGGGTAGTACAAAAGGGAAATGTTGGTGGGAGCCCCAC 988
DB 38094 CCTATGAGAGAACCTCTCCAGGGTAGTACAAAAGGGAAATGTTGGTGGGAGCCCCAC 38153
QY 989 ACAGATCCCCAGTGGGCTCCATCTAGTAGAGCTGTGAGAGAAAGTCCACCATCTCCA 1048
DB 38154 ACAGATCCCCAGTGGGCTCCATCTAGTAGAGCTGTGAGAGAAAGTCCACCATCTCCA 38213
QY 1049 GACTCCAGAAGGATAGATCCACTGACAGCTTGCAGCATGTGCTGCTGAAAAATCCAGACA 1108
DB 38214 GACTCCAGAAGGATAGATCCACTGACAGCTTGCAGCATGTGCTGCTGAAAAATCCAGACA 38273
QY 1109 CTCAGTGCACCGCTGTGAAAGCAGCAGGATGGAGTCTGTACCCCTACAAAACCGTAGTG 1168
DB 38274 CTCAGTGCACCGCTGTGAAAGCAGCAGGATGGAGTCTGTACCCCTACAAAACCGTAGTG 38333
QY 1169 CAGAGCTCACCAGACCGTGGGAATCTACCTCTTCATGTCATGACCTGGACGTGAGAC 1228
DB 38334 CAGAGCTCACCAGACCGTGGGAATCTACCTCTTCATGTCATGACCTGGACGTGAGAC 38393
QY 1229 ATGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTTGACTGCCCCCTGCGATTTCGA 1288
DB 38394 ATGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTTGACTGCCCCCTGCGATTTCGA 38453
QY 1289 CTTATATGGGGCCGTCACCCCTTTGTTGGCCAATTTTTCATTTGGAAGTCCCGTAT 1348
DB 38454 CTTATATGGGGCCGTCACCCCTTTGTTGGCCAATTTTTCATTTTGAAGTCCCGTAT 38513
QY 1349 TTACCCATGCTGCTACCTCCATGCTATG 1377
DB 38514 TTACCCATGCTGCTACCTCCATGCTATG 38542
```

RESULT 2  
AC017063/c

LOCUS AC017063 193092 bp DNA linear PRI 27-MAR-2002  
DEFINITION Homo sapiens BAC clone RP11-354H17 from 4, complete sequence.

ACCESSION AC017063

VERSION AC017063.8 GI:19482379

KEYWORDS HTG.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
1 (bases 1 to 193092)
Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
9847074
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
2 (bases 1 to 193092)
Boatright,E., Haakenson,W. and Doebber,A.
The sequence of Homo sapiens BAC clone RP11-354H17
Unpublished (2001)
3 (bases 1 to 193092)
Waterston,R.H.
Direct Submission
Submitted (09-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 193092)
Waterston,R.H.
Direct Submission
Submitted (15-MAR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 193092)
Waterston,R.
Direct Submission
Submitted (27-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 15, 2002 this sequence version replaced gi:16596637.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
-----
Summary Statistics
-----
Center project name: H_NH0354H17
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Oseegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-756P10, 2000 bp overlap;
the clone sequenced to the right is RP11-818C3. Actual start of
this clone is at base position 178381 of RP11-756P10; actual end is
at base position 193092 of RP11-354H17.

There is an unresolved tandem repeat from 23045 to 24207.

```

---

```

FEATURES
source
Location/Qualifiers
1..193092
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="UL"
/map="4"
/clone="RP11-354H17"
/clone_lib="RPCI-11"
467..619
/rpt_family="L1"
840..880
/rpt_family="TG)n"
881..965
/rpt_family="(TA)n"
984..1039
/rpt_family="(TA)n"
1040..1071
/rpt_family="(TTTA)n"
1072..1210
/rpt_family="Alu"
1211..1478
/rpt_family="Alu"
1479..1490
/rpt_family="Alu"
2394..2715
/rpt_family="MaLR"
3180..3470
/rpt_family="Alu"
3891..3996
/rpt_family="MIR"
4022..4053
/rpt_family="AT_rich"
4291..4417
/rpt_family="MIR"
4830..4999
/rpt_family="MER1_type"
6479..6507
/rpt_family="AT_rich"
6691..6989
/rpt_family="Alu"
7734..7843
/rpt_family="L1"
9245..9274
/rpt_family="AT_rich"
10159..10408
/rpt_family="MIR"
10793..10968
/rpt_family="(TATAA)n"
10982..11144
/rpt_family="(TA)n"
11157..11193
/rpt_family="(TG)n"
12999..13157
/rpt_family="L2"
13192..13238
/rpt_family="AT_rich"
13297..14044
/rpt_family="L1"
14366..14429
/rpt_family="MIR"
14565..15340
/rpt_family="L1"
15608..16083
/rpt_family="MaLR"
16094..16267
/rpt_family="MIR"
17080..17178
/rpt_family="GA-rich"
17211..17235
/rpt_family="AT_rich"
17237..17522
/rpt_family="Alu"
17581..17975

```

```
repeat_region /rpt_family="L1" 18061..18371
repeat_region /rpt_family="L1" 20140..20312
repeat_region /rpt_family="MER1_type?" 20468..21181
repeat_region /rpt_family="L2" 21208..21485
repeat_region /rpt_family="MER1_type" 21488..21697
repeat_region /rpt_family="L2" 22183..22404
repeat_region /rpt_family="ERVL" 22519..22567
repeat_region /rpt_family="(CA)n" 22636..22815
repeat_region /rpt_family="(CA)n" 22877..23053
repeat_region /rpt_family="(CA)n" 23075..23388
repeat_region /rpt_family="(CA)n" 23351..23528
repeat_region /rpt_family="(CAGA)n" 23595..23832
repeat_region /rpt_family="(CAGC)n" 23822..24006
repeat_region /rpt_family="(CA)n" 24018..24148
repeat_region /rpt_family="(CA)n" 25078..25359
repeat_region /rpt_family="L1" 27832..27895
repeat_region /rpt_family="L1" 27922..27955
repeat_region /rpt_family="(TG)n" 27967..27995
repeat_region /rpt_family="(GA)n" 28974..29261
repeat_region /rpt_family="Alu" 29271..29301
repeat_region /rpt_family="(TAA)n" 30597..30777
repeat_region /rpt_family="L1" 30833..30908
repeat_region /rpt_family="(TATATG)n"

Query Match 50.5%; Score 695.2; DB 9; Length 193092;
Best Local Similarity 80.5%; Pred. No. 5.8e-195;
Matches 853; Conservative 0; Mismatches 173; Indels 34; Gaps 2;

Qy 351 GAATATTGGCTAATAGGAAGCAAAATTTGGCTAATAAGGAAGAAATTTCTAAGCAGCA 410
Db 113937 GAATTTTAAGTTGGGAGAGATGATTTAGGGTATCTGGCAGATAAATTTCTAAGCAGCA 113878

Qy 411 AAGCATTTCAAGAGTGACTTGGTGGCTGTTAAAGGCATTTCAGTTTCATAAGGGAGGCAGA 470
Db 113877 AAGCATTTCAAGAGTGACTTGGTGGCTGTTAAAGGCATTTCAGTTTCATAAGGGAGGCAGA 113818

Qy 471 GCATAAGATTCAGAAAATTTGCACCCCTGACATGTCATAAAGAAAACCCATTTC 530
Db 113817 GCATAAATAATTTGGAAAATTTGCAGCCTGACATGTCATAAAGAAAACCCATTTC 113758

Qy 531 TGAGGGAAATTCAGCTGGCTGCAGAAAATTTGCATATGTAATGAGGAGCTGAATGTAA 590
Db 113757 TGAGGAGAAATTTGAGCTGGCTGCAGAAAATTTACATAGTAATGAGGAGCTGAATGTAA 113698

Qy 591 TCCTCAAGACAATGGGAAAATATTCCTCGACATGTCAGAGGTCCTTCACAGCAGTCAT 650
Db 113697 TCCCAAGACAATGGGAAAATGCTCCAAAGCATGTCAGAGGTCCTTAACAGCAGCCCT 113638

Qy 651 CAATCACTGGCTGGAGGCTAGGAGAAAATGGTTTTTGGGACAGGCCAGGGTCCCT 710
Db 113637 ACCATCAGGCTGGAAGGCTAGGAGAAAATGGTTTTTGGGACAGGCCAGGGTCCCT 113578
```

```
Qy 711 GTGCTGTGTGAGCCTAGAGACTTGGTCCCTGTGTGCCAGTTAATTCA----- 759
Db 113577 GTGCTGTGTGAGTCTAGGTACTTGGTCCCTGTGCATCTAGCCACTCCAGACATGACTAA 113518
Qy 760 -----GCTGTGCTTCAGAGGGTCAAGCCCCCAAGCCCTTGCGA 797
Db 113517 AAGGGGCAAGGTACAGCTCAGGCTGTGCTTCAGAGGGTGAAGCTGCAAGCCCTTGCG 113458
Qy 798 GCTTCCAAAGTGTGTGAGCCTGTGGGTGCAAGAAGTCAAGAATTGAGGTTTGGGAACC 857
Db 113457 GCTTCCAAATGTTGAACCTGCGGTTCACAGAAATCAAGAATTGAGTTTGGGAACC 113398
Qy 858 TCCAATCAGATTTCAGAAATATATGGAACCCCTGGATGCCAGGCAGAAATTTGCTGT 917
Db 113397 TCCGCTAGATTTCAGAAAATGTATGAAAATGCTTGGATGCCAGGCAGAAATTTGCTGTC 113338
Qy 918 AGGGGTGGGTCTCATGGAGAACTCTGCAAGGCTAGTACAAAAGGGAATTTGGGTG 977
Db 113337 AGGAGTGGGGCCCTCATGGAGAACTCTGCTAGGSCAGTCAGAAAGGGAATTTGAGGTG 113278
Qy 978 GGAGCCCCACACAGAGTCCCGCAGTGGGGCTCCATCTAGTAGAGCTGTGAGAAGAAGTCC 1037
Db 113277 AGAGCCCCACACAGAGCCCTCTGCGGACACCACTAGTGGAGCTGTGAAAAGAGGAC 113218
Qy 1038 ACCATCTCTCCAGACTCCAGAAAGGCTAGATCCACTGCACAGCTTGCAGCATGTGCCCTGAAAA 1097
Db 113217 ACTGCTCTCCAGACCCCAAGATGTTAGATCCACTGCAGAGCTTGCACCATGCTTGGAAA 113158
Qy 1098 ATCCACAGACACTCAGTGCCAGCCTGTGAAAGCAGCAGGGATGGAGTCTGTACCCCTACAA 1157
Db 113157 AGCCACAGACACTGAACACACAGACCATGAAAGCAGCCAGAGGGTGGCTGTACCCCTACAA 113098
Qy 1158 AACCGTAGTGCAGAGCTGACCAAGACCGTGGGAATCTACCTTTCATTTGTCATGACCT 1217
Db 113097 AGCCACAGGGTGAAATTTGGCCAAAGACCAAGCAAGAACCCACCTCTTGCATCAGCATGACT 113038
Qy 1218 GGAGCTGAGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTTGACTGCCCCAC 1277
Db 113037 GGATGTGAGACATGTTGTCAAAGGACATTAATTTGGAGCTTTAAGATTTGACTGCCCCAC 112978
Qy 1278 TGGATTTCCGACTTATATGGGGCCCGT-ACCCCTTTGTTTGGCCAAATTTTTCATTTG 1336
Db 112977 TGGATTTTGGACTTGCACGGGGCTGTAAACCCCTTTGTTTGGCCAAATGCTCTCATTTG 112918
Qy 1337 GAACCTGCGTATTTACCAATGCGCTGTACCTCCATTTGAT 1376
Db 112917 GATTTGCTGTTATTTATCCAATGCGCTGTACTGCCAATGTTT 112878
```

```
RESULT 3
AL138962/c
LOCUS Human DNA sequence from clone RP11-279N8 on chromosome 13, complete
DEFINITION sequence.
ACCESSION AL138962
VERSION AL138962.16 GI:12597036
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Plimates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 124518)
Bates,K.
Direct Submission
Submitted (28-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jan 29, 2001 this sequence version replaced gi:12584354.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
```

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human

chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr13>

RP11-279N8 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone

RP11-279N8 It may be shorter because we sequence overlapping

sections only once, except for a 100 base overlap.

The true right end of clone RP11-279N8 is at 124518 in this

sequence. The true left end of clone RP11-528024 is at 114969 in

this sequence. The true right end of clone RP11-214011 is at 100 in

this sequence.

FEATURES

source 1. 124518

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="13"

/clone="RP11-279N8"

/clone.lib="RPCI-11.1"

/note="Big dye primer and dGTP big dye chemistries only."

misc\_feature 112388..112496

BASE COUNT 39295 a 26385 c 24794 g 34044 t

ORIGIN

Query Match 50.2%; Score 691; DB 9; Length 124518;

Best Local Similarity 81.8%; Pred. No. 9.9e-194;

Matches 837; Conservative 0; Mismatches 150; Indels 36; Gaps 2;

QY	390	GAAGAATAATCTTAAGCAGCAAGAAATTCAGAGGTGACTTGGGTGCTGTTAAAGGCATT	449
DB	66716	GAAGACATTTCTAAGCAGAAAGCAATCAAGAGGTGACTTGGGTGCTGTTAGAGGCATC	66657
QY	450	CAGTTTCATAGGGAGGAGCAGACATTAAGATTTCAGAAATTTGCACCTCGACATGTGAT	509
DB	66656	CAGTTTATAAGGGAACAGAGATATAAAGTTTGAAAATTTGCAGCCTGACAAATGTGAT	66597
QY	510	AAAAAAGAAAACCCATTCTTGAGGGGAAATTCAGCTGGCTGCAGAAATTTGCATATG	569
DB	66596	AGAAAAGAAATCCATTTCTGAGAGCAAGACTCAAGCTGGCTGCAGAAATTTGCATAG	66537
QY	570	TAATGAGGAGCTGAATGTTAATCTCAAGACAAATGGGAAAATATCTCTCGACATGTCA	629
DB	66536	TAACGAGGAGCCAAATGTTAATCAACCAAGACAAATGAGGAAATGCTCCAGGGCATGTCA	66477
QY	630	GAGGTCTTCACAGCAGTTCATCAATCACTAGCCTGGAGGCGCTAGCAG---AAATGGTT	686
DB	66476	GAGGTTCATCACAGCATCCCTCCCATCACAGCCTTGAAGCTTAGGAGGAAAAATGGTT	66417
QY	687	TTGTGGGACAGCCCGAGGCTCCTGCTGTGTGTCAGCTGAGACTTGGTCCCTGTGT	746
DB	66416	TCATGGAAGGGGCACAGGCTCCTGTGTGTGTCAGCTAGGACTTGGTCTCTGCAT	66357
QY	747	CCAGTTTAATTC-----GCTGTGGCTTCAGA	773
DB	66356	CCAGCTGCTCCAGCTGTGGCTGAAAGGGGCCAATGTGAGTTTCAGGCTGTGGCTTCAGA	66297
QY	774	GGGTGCAGCCCCCAAGCCTTGGCAGCTTCCAAAGTGGTTGAGCCTGTGGGTGCAAGAA	833

```
Dye-terminator Big Dye: 86% of reads
Consensus quality: 76604 bases at least Q40
Consensus quality: 77098 bases at least Q30
Consensus quality: 77320 bases at least Q20
Insert size: 77405; sum-of-contigs
Insert size: 85298; 1.5% error; agarose-fp
Quality coverage: 4.54x in Q20 bases; sum-of-contigs Quality
coverage: 4.30x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
    source
        Location/Qualifiers
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="X"
            /map="q21.33-22.3"
            /clone="RP11-435A2"
            /clone_lib="RPC1-11.2"
            misc_feature
                1..77405
                    /note="assembly fragment:00554"
BASE COUNT      25072 a 14551 c 15133 g 22649 t
ORIGIN
Query Match      50.2%; Score 690.8; DB 2; Length 77405;
Best Local Similarity 80.6%; Pred. No. 1.1e-193;
Matches 871; Conservative 0; Mismatches 172; Indels 37; Gaps 4;

Qy 334 CTGCATCATGCCAAGGTGAATATTTGGCTAATGAGGAAGCAAAATTTGGCTAATAAGGAA 393
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8637 CTGCCCTAGAGATCTGGGATTTGACCTTGAGAGAGATGATTTAGGCTATCAGGTGGAA 8696

Qy 394 GAAATTTCTAAGCAGCAAGACATTCAGAGGTGACTTGGGTGCTTTAAAGGCATTCAGT 453
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8697 GAAATTTCTAAGCAGCAAGACATTCAGAGGTGACTTGGGTGCTTTAAAGGCATTCAGT 8756

Qy 454 TTCTAAGGGAGGAGCAGCATTAAGAGTTTCAGAAAATTTGCACCTGCACAATGTGATAAAA 513
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8757 TTTGCAAGGGAAGCAGAACATAAAGTTTCAGAAAATTTTCAGCTGCACAATGTGATAAA 8816

Qy 514 AAGAAAACCCA-TTTTCTGAGGGGAATTCACAGCTGGCTGCAGAAATTTGCATATGTAA 572
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8817 AAGAAAACCCCATTTTCTGATGATAAATTCAGCCAGCTGCAGAAATTTGCATAAGTAA 8876

Qy 573 TGAGGAGCTGAATGTTAATCTCAAGACAATGGGGAATAATCTCCCTGGACATCTCAGAG 632
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8877 CGAGAGCCCAATGTTAATCACTAAGACCATGGGGAATAATCTCCATGGCTTGTTCAGAG 8936

Qy 633 GTCTTCACAGCAGTCCCAATCACTGG--CCTGGAGGCCCTAGGAGAAAATGGTTTTGT 690
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8937 GTCTTCACGCGAGCCCTCACATCACTGCGCCCCAGAGGCCCTAGGAGAAAATGCTTTAT 8996

Qy 691 GGGACAGGCCAGGCTCCCTGTGTGTGTGAGAGCTAGAGACTTGGTGCCTGTGTGCCA 750
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8997 GGGTCGGGCCAGGCTCCCATGCTATGTGCAGTCTAGGAGCTTGGTGCCTGCTGCCA 9056

Qy 751 GTTAATTCAC-----GCTGTGGCTTCAGAGGCT 777
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 9057 GCCACTCCAGCCCACTAAGAAAGGGGCCAAGGTATAGTTGGGCTGTGTCTTCRAAGGAT 9116

Qy 778 GCAAGCCCCAAGCCTTGGCAGCTTCCCAAGTGGTGTGAGCCTGTGGGTGCAAGAAGTCA 837
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 9117 GGAAGCCCCAGGCTTGGCAGCTTCCATGTGTGTGTGAGCCTGCAGGTGCACCTGAAGTCA 9176

Qy 938 AGAATTCAGGTTTGGGAACCTCCCAATCAGATTTCAGAAGATATATGGAACCCCTGGATG 897
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 9177 AGAATTCAGGTTTGGGAACCTCCCAATCAGATTTCAGAAGATATATGGAACCTGGATG 9236

Qy 998 CCCAGGCGAAGTTTGTGTAGGGGTGGGTCCCTCATGGAACCTCTGCAAGGGTACTA 957
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 9237 CCCAGGCAAAAGTTTGTGTGCGAGGGCGAGGACCTCATGGAAGAACCTCTGTAGGGTGGTG 9296
```

```
repeat_region 8954..9088
/note="L1 element fragment"
repeat_region 12171..13985
/note="L1 element fragment"
repeat_region 14179..14612
/note="L1 element fragment"
repeat_region 14662..14780
/note="L1 element fragment"
repeat_region 14824..15044
/note="L1 element fragment"
repeat_region 15234..15369
/note="L1 element fragment"
repeat_region 15421..15849
/note="L1 element fragment"
repeat_region 16026..16155
/note="MSTA element fragment"
repeat_region 16073..16149
/note="THE1B element fragment"
repeat_region 16204..16495
/note="Alu repeat: matches 1..308 of consensus"
repeat_region 16532..16609
/note="THE1B element fragment"
repeat_region 16624..16722
/note="MSTC element fragment"
repeat_region 16626..16737
/note="MSTA element fragment"
repeat_region 16651..16712
/note="THE1B element fragment"
repeat_region 17060..17653
/note="L1 element fragment"
repeat_region 19720..19797
/note="MER21B element fragment"
repeat_region 19877..19915
/note="MER21B element fragment"
repeat_region 20007..20256
/note="MER21B element fragment"
repeat_region 20518..20734
/note="MER30 element fragment"
repeat_region 22661..22730
/note="MLT1B element fragment"
repeat_region 22740..22814
/note="MLT1A element fragment"
repeat_region 22917..22974
/note="MLT1A element fragment"
repeat_region 23342..23501
/note="THE1B element fragment"
repeat_region 23550..23697
/note="THE1B element fragment"
repeat_region 23725..24279
/note="THR element fragment"
repeat_region 24395..24593
/note="THR element fragment"
repeat_region 24643..25270
/note="THR element fragment"
repeat_region 25277..25463
/note="THE1B element fragment"
repeat_region 25528..25627
/note="THE1B element fragment"
repeat_region 26081..26196
/note="MLT1A element fragment"
repeat_region 26265..26661
/note="LTR2 element fragment"
repeat_region 26797..26898
/note="MSTC element fragment"
repeat_region 27949..28241
/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 28319..28608
/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 30032..30159
/note="MERS element fragment"
repeat_region 31794..31980
```

---

```
/partial
/note="Alu repeat: matches 308..109 of consensus"
31981..32073
/partial
/note="Alu repeat: matches 93..1 of consensus"
32232..32735
/note="MER9 element fragment"
35168..35654
/note="MER4B element fragment"
40640..40693
/note="27 copies of 2 mer 94 & conserved"

BASE COUNT 11544 a 7879 c 8141 g 13150 t
ORIGIN
Query Match 50.1%; Score 690.2; DB 9; Length 40714;
Best Local Similarity 80.6%; Pred No. 1.5e-193;
Matches 870; Conservative 0; Mismatches 173; Indels 36; Gaps 4;

QY 334 CTGCATCAATTGCAAGTGAATATTTGGCTAATGAGGAACAAATTTTGCTTAATAAGGAA 393
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24003 CTGCCCTAGAGATCTGTGGAATTTGACCTTGAGAGAGATGATTTAGGCTATCAGGTGAA 24062

QY 394 GAAATTTCTAACGACGAAGCAATTCAGAGGTGACTTGGTGCTGTTAAAGGCATTCAGT 453
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24063 GAAATTTCTAACGACGAAGCAATTCAGAGGTGACTTGGTGCTGTTAAAGGCATTCAGT 24122

QY 454 TTTCATAAGGGAGGAGAGCATTAAGAGTTTCAGAAAAATTTGCACCCCTGACAAATGTGATAAAA 513
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24123 TTTGCAAGGGAAGCAGACATAAAGTTTCAGAAAAATTTTCAGCCCTGCAAAATGTGATAAA 24182

QY 514 AAGAAAAACCCA-TTTTCTGAGGGGAAATTCAGCTGGCTGCAGAAATTTTGCATATGTAA 572
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24183 AAGAAAAACCCAATTTTCTGATGATAAATTTGAAGCAGCTGCAGAAATTTTGCATAAGTAA 24242

QY 573 TGAGGAGCTGAATGTTAATCTCAAGACAATGGGGAATATATCTCTCGACATGTCAGAG 632
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24243 CGAGGAGCCAAATGTTAATCACTAAGACCATGGGAAAAATATCTCCATGGCTTGTGAGAG 24302

QY 633 GTCCTTACAGCAGCTCCATCAAAATCACTG-GCCTGGAGGCTAGGAGAAAAATGTTTGTG 691
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24303 GTCCTTACAGGAGCCCTCACAATCACTGCCCCAGAGGCTAGGAGAAAAATGTTTATG 24362

QY 692 GGACAGGCCAGGGTCCCTGTGTGTGTGACAGCTAGAGACTTGGTGCCTGTGCTCCAG 751
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24363 GGTGCGGCCAGGGTCCCATGCTATGTCAGTCTAGGAGCTTGGTGCCTGTGCTCCAG 24422

QY 752 TTAATTCA-----GCTGTGGCTTCAGAGGGTG 778
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24423 CCACCTCCAGCCATAACTGAAAGGGGCCAAGGTATAGCTTGGGCTGTTGCTTCAAGAGATG 24482

QY 779 CAAGCCCCAAGCCTTGGCAGCTTCCAAGTGGTGTTCAGAGCTGTGGGTGCAAGAGCTCAA 838
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24483 GAAGCCCCAGGCTTGGCAGCTTCCATGTGGTGTGGAGCTTCGAGGTGCACCTGAACTCAA 24542

QY 839 GAATTGAGTGTGGAAACCTCCAATCAGATTTTCAGAGATATATGGAACCCCTGGATGC 898
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24543 GAATTGAGTGTGGAAACCTCCACCTAGATTTTCAGAGATGATGGAATGCTGCTGATGC 24602

QY 899 CCAGGCAAGAGTTTGTGTAGGGGTGGGTCTCTCATGGAGAACCTCTGCAAGGGTAGTAC 958
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24603 CCAGCAAAAGTTTGTGTAGGGGTGGGTCTCTCATGGAGAACCTCTGTTAGGGTGGTGT 24662

QY 959 AAAAGGGAATGTGGTGGGAGCCCCACACAGAGTCCCCAGTGGGGCTCATCTAGTA 1018
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24663 GGAAGGGAAGAGTGGGGTCTATAGCCCCACACAGAGTCTCTACCAGGGACCACTAGTG 24722

QY 1019 GAGCTGTGAGAAAGAGTCCACCATCTCCAGACTCCAGAGGGTAGATCCACTGACAGT 1078
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24723 GAGCTGTGAGAAAGAGGGCCACCATCTCTCCAGAGCCCCAGAAATGGTAGATCCACGACAGCC 24782

QY 1079 TCCAGCATGTGCTGTAATAATCCACAGACACTCAGTGGCCAGCTGTGAAAGCAGAGGGA 1138
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24783 TGCACCGTGTGCTTGGAAAAAGCCGACAGACTCAATGTCAGGCTGTGTAAGAGCAGAGGGA 24842
```

Qy 1139 TGGAGTCTGTACCCCTACAAAACCGTAGTGGCAGAGCTGACCAAGACCGGTGGGAATCTACC 1198  
Dy 24843 GGGAGGCTGTACCCCTGCAAGACCCAGGGCGAGAGCTGCCCAAGATCATGTGGAACCAATC 24902  
Qy 1199 TCTTGCAATGTCAGACCTGACGCTGAGACATGAGTCAAAAGAGATCATTTTGGAGCTT 1258  
Dy 24903 TCTTGCAATGTCAGACCTGAGATGAGACATGAGTCAAAAGAGATCATTTTGGAGCTT 24962  
Qy 1259 TAAGATTTGACTGCCCCACTGGATTTGGAGCTTATATGGGCCCGTA-CCCTCTTTGTTTT 1317  
Dy 24963 TAAGATTTGACTGCCCCACTGGATTTGGAGCTTATATGGGCCCGTGTAGCCCTTTGTTTT 25022  
Qy 1318 GGCAATTTTTCATTTGGAACTGCGGTATTTATCCCAATCCCTGTACCTCCATTTGAT 1376  
Dy 25023 GGCAATATCTCCCATTTGGAACTGCGGTATTTATCCCAATCCCTGTACCTCCATTTGAT 25081

RESULT 6  
AC008799 123098 bp DNA linear PRI 28-MAR-2002  
LOCUS Homo sapiens chromosome 5 clone CTD-2061E19, complete sequence.  
DEFINITION AC008799  
AC008799.6 GI:19774451  
VERSION HTG.  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 123098)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 123098)  
DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (07-MAR-2002) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 123098)  
DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (07-MAR-2002) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
4 (bases 1 to 123098)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (28-MAR-2002) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
On Mar 28, 2002 this sequence version replaced gi:19224702.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.8% of Sequence;  
Estimated Total Number of Errors is 0.4.  
FEATURES  
source  
1. 123098  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-2061E19"  
BASE COUNT 37998 a 24397 c 23979 g 36724 t  
ORIGIN

Query Match 50.08; Score 689; DB 9; Length 123098;  
Best Local Similarity 80.11; Pred. No. 3.9e-193;  
Matches 850; Conservative 0; Mismatches 175; Indels 36; Gaps 2;

Qy 351 GAATATTTGGCTTAATGAGCAAGCAAAATTTTGGCTTAATAAGAGAAATTTCTAAGCAGCA 410  
Dy 107797 GAATTTTAACCTTGAAGAAATGATTTAGGGTATCTGCGCAAGAAATTTCTAAGCAGCA 107856  
Qy 411 AAGCATTCAAGAGGTGACTTGGGTGCTGTTAAAGGCATTCAGTTTCAAGGGAGGCAGA 470

Dy 107857 AAGCATTCCAGAGGTGACTTGGGTGCTGTTAAAGGTATTTCAGTTTATAAGGAAGCAGA 107916  
Qy 471 GCATAAGAGTTTCAGAAAATTTGCACCCCTGCACAACTGTATAAAGAAAGAAACCCATTTC 530  
Dy 107917 GCATAAAGAGTTGGAATAATTTGACAGTCCGACAAATGTATAGAAAAGAAATTTCCATTTC 107976  
Qy 531 TGAGGGGAAATTCAGCTGGCTGCAGAAATTTGCATATATGTAATGAGGAGCTGAATGTTAA 590  
Dy 107977 TGAGGAGAAATTCAGCTGGCTGCAGAAATGTGCATTAAGTACAGAGCTGAATATTA 108036  
Qy 591 TCCTCAAGACAATGGGAAAATATCTCTCGACATGTGCAGAGTCTTCACAGCAGTCCAT 650  
Dy 108037 TCCCAAGACAATGGGAAAACATCTCCAAGGCATGTTCAGAGGTCTTCACGCGACCCCT 108096  
Qy 651 CAATCACTGCGCTGGAGGCTAGGAGAAA- - -TGGTTTGTGGGACAGGCCAGGGTC 707  
Dy 108097 CCCATCAAAACCCNAAGCCCTAGGAGGAAAAGTGGTTTGTGGGCCAGGCCAGGGTC 108156  
Qy 708 CTGTGCTGTGTCAGCCTAGAGACTTGTGCTCTGTGCCAGTTAATTC- - - - - 759  
Dy 108157 CTGTGCTGTGTCAGCTTAGGAGCTTGTGCTCTGCATCCAGCTGTCCAGCCATGCG 108216  
Qy 760 - - - - -GCTGTGCTTTCAGAGGTGCAAGGCCCAAGCCCTTG 794  
Dy 108217 TGAAGGGGCCAACATAGAGCTCGGGCTGTGGCTTTCAGAAAGTGCAGCCCAAGCCCTG 108276  
Qy 795 GCAGCTTCCAAAGTGTGTGAGCCCTGTGGTGCAGAAAGTCAACAATTTGAGGTTTGGGA 854  
Dy 108277 GCAGCTTCCAAAGTGTGTGAGCCCTGCAAGTGCAGAAAGTCAACAATTTGAGGTTTGGGA 108336  
Qy 855 ACCTCCAATCAGATTTCAAGAGATATATGGAACCCCTGGATGCCAGCAGCAGAAATTTGCG 914  
Dy 108337 ACCTCCACCTAGATTGCAGATGATGTAAGAAATGCCCGGATGACCCAGCAGAAATTTGCG 108396  
Qy 915 TGTAGGGTGGGTCTCTCATGAGAAACCTCTGCAGAGGTAGTACAAAGGGGAAATTTGCG 974  
Dy 108397 TGCAGGGGCGAGGCGCTTTCAGAGAAACCTCTGCTAGGCGAGTGCAGAGAAATTTGAGG 108456  
Qy 975 GTGGGAGCCCCACACAGAGTCCCGAGTGGGCTCCATCTAGTAGAGTGTGTCAGAGGAAG 1034  
Dy 108457 GTCAGAGGCCACACACAGAGTACCCACTAGAGCACTGCTAGTGGAGCTGTGAGAGGAAG 108516  
Qy 1035 TCCACCATCTCCAGACTCCAGAGGGTAGATPCCACTCAGACGTTGCGACGATGTGCCCTGA 1094  
Dy 108517 GCCACCATCTCCAGACTCCAGAGGGTAGTGGTCCAGCAGCTGTGCTGTACCTGG 108576  
Qy 1095 AAATCCACACACACTAGTGGCCAGCTGTGAAACAGCAGGGGATGGAGTCTGTACCTTA 1154  
Dy 108577 AAAAGCCACAGGCACACTCATTTGCCAGCTGTGAAAGCAGCCAGGAGGAGGCTATCCCTG 108636  
Qy 1155 CAAACCCGTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTCTTTCAGTTGTCATGA 1214  
Dy 108637 CAAAGTCAGGGGGGGAAGCTGTCCAGACCTTTGGGAACCTACCTCTTTCAGCTCAGCTGA 108696  
Qy 1215 CTGGAGCTGAGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTTGACTGCC 1274  
Dy 108697 CCCGATCTGAGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTTGACTGCC 108756  
Qy 1275 CACTGGATTTCCGACTTATATGGGGCCCGTACCCCTTTGTTTTGGCCAAATTTTTCATT 1334  
Dy 108757 TGCTGATTTTCAGACTGTCATGGGCTGTAGCCCTTGTGCTTTGGCCAAATTTTTCATT 108816  
Qy 1335 TGGAACTCCGCTATTTACCCCAATGCTGTACCTCCATTGTA 1375  
Dy 108817 TGGAACTGCTATTTACCCCAATGCTGTACCCCAATGTA 108857

## RESULT 7

AC025577

LOCUS

DEFINITION

AC025577 154090 bp DNA linear PRI 30-APR-2002  
Homo sapiens 12 BAC RP11-13C3 (Roswell park Cancer Institute Human  
BAC Library) complete sequence.

AC025577  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

AC025577.15 Gi:9910028  
 HTG.  
 Homo sapiens.  
 Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 154090)

REFERENCE  
 AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,  
 Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
 Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,  
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
 Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,  
 Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
 Dugan-Rocha,S., Durbin,K.J., Eamhart,C., Edgar,D., Edwards,C.C.,  
 Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,  
 Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,  
 Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,  
 Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,  
 Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,  
 Hodgson,A., Hognes,M., Hollway,C., Hollins,B., Homsi,F.,  
 Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
 Karlsson,E., Kelly,S., Khan,D., King,L., Korvah,J., Kovar,C.,  
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,  
 Lewis,L., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
 Louisgied,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
 Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,  
 Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P.,  
 Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,  
 Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,  
 Morris,S., Moser,M., Muzny,D., Neal,D., Nelson,D., Newton,J.,  
 Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E.,  
 Nwokwenko,S., Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A.,  
 Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E.,  
 Pu,L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojebokan,I.,  
 Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H.,  
 Shm,C., Shooshitari,N., Sisson,I., Sodergren,E., Sonaikhe,T.,  
 Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P.,  
 Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T.,  
 Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V.,  
 Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,  
 Washington,C., Watlington,S., Williams,G., Williamson,A.,  
 Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,  
 Zorilla,S., Kucherlapati,R., Nelson,D. and Gibbs,R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 154090)  
 Worley,K.C.  
 Direct Submission  
 Submitted (11-MAR-2000) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 154090)  
 Worley,K.C.  
 Direct Submission  
 Submitted (25-AUG-2000) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 4 (bases 1 to 154090)  
 Worley,K.C.  
 Direct Submission  
 Submitted (30-APR-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Aug 25, 2000 this sequence version replaced gi:9664948.  
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
 gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the  
 entire insert of this clone. Overlapping regions of clones are only  
 sequenced and submitted once, so the sequence for the remainder of  
 the insert may be found in the record for the adjacent clones.  
 Overlapping clones are noted at the beginning and end of the  
 Features listing.

# ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches  
 of a local database that includes entries from dbSTS, GDB, and  
 local mapping efforts.  
 Repeats are identified using RepeatMasker (A. Smit and P. Green,  
 unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
 EST and cDNA sequences. Genes demonstrate at least two exons  
 flanked by consensus splice sites that maintained sequence  
 continuity across the splice junctions. Sequences that are not  
 identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
 standard of double strand coverage with a minimum of 2 clones and 2  
 reads with no ambiguities or 2 chemistries with a minimum of 2  
 clones and 3 reads with no ambiguities. If the sequence quality for  
 a region does not meet this standard, it will be indicated in the  
 annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
 standards - estimated error rate less than 1 per 10,000 bases.  
 Reports of low quality individual bases and measures of base  
 quality are listed below. Description of the metrics can be found  
 at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

## QUALSTAT-REPORT-----

----- Summary Statistics -----  
 Contig length: 154090  
 Phrap values in estimate: 153545  
 Average error rate (BCM-Phrap estimate): 0.00261755  
 Fraction of Phrap values less than 40 : 0.0284086  
 Number of consensus changing edits: 22  
 Number of N's in consensus : 0

----- Consensus changing edits -----  

Position	Original+Context	Edited+Context
9347	aacacagacc(n)ttttttttt	aacacagacc(t)ttttttttt
10135	attgcccct(n)taggaaaaga	attgcccct(t)taggaaaaga
10353	ctgtcacccag(n)aaagatcca	ctgtcacccag(t)aaagatcca
10417	agcaatgcag(n)ggctacagaa	agcaatgcag(t)ggctacagaa
61959	ctttactatt(n)tgtttatct	ctttactatt(t)tgtttatct
61960	tttactattn(t)gtttattcta	tttactattn(g)gtttattcta
61961	ttactattnt(g)tttattctaa	ttactattnt(t)tttattctaa
61982	caacacaaag(n)ntaggtttgg	caacacaaag(a)ntaggtttgg
61983	caacacaaag(n)taggtttggn	caacacaaag(t)taggtttggn
61993	ntaggtttgg(n)ttactcttcc	ntaggtttgg(t)ttactcttcc
62205	tcattgacct(n)ctggctattc	tcattgacct(a)ctggctattc
62956	actgcaacct(n)tgctcccag	actgcaacct(c)tgctcccag
62972	ttcacatata(n)cagtcacagta	ttcacatata(a)cagtcacagta
90212	cctagaaaaa(n)gacntttctt	cctagaaaaa(t)gacntttctt
90216	gaaaaangac(n)tttctttttt	gaaaaangac(t)tttctttttt
90228	tttcttttta(n)tttttctnat	tttcttttta(t)tttttctnat
90236	taattttttc(n)atcttttaggg	taattttttc(t)atcttttaggg
90443	accttagatg(n)ctctcccag	accttagatg(a)ctctcccag
91405	cacagcttaa(n)caaggttaaga	cacagcttaa(a)caaggttaaga
122290	aaaaaagaaa(n)gtaagagaaa	aaaaaagaaa(t)gtaagagaaa
137090	acagagaaaag(n)caaaacactc	acagagaaaag(a)caaaacactc
141632	tgaagccact(n)gaaaaagtaat	tgaagccact(t)gaaaaagtaat

----- Distribution of Quality < 40 Bases -----

```
1000| * * *
900| * * *
800| * * *
700| * * *
600| * * *
500| * * *
400| * * *
300| * * *
200| * * *
100| * * *
0| * * *
-----
5 10 15 20 25 30 35 40
Phrap Value Range
```

```
-----
Version: 1.01 qxfo.
FEATURES
  source
    Location/Qualifiers
      1..154090
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="12"
        /clone="RP11-13C3"
        1..1231
        /note="overlaps bases 14771..16001 of clone AC079952"
        /function="clone overlap"
        54..288
        /rpt_family="MIR"
        1020..1146
        /rpt_family="MER3"
        complement(1248..1584)
        /rpt_family="MER46B"
        complement(1624..1682)
        /rpt_family="MIR"
        2630..2761
        /rpt_family="GA-rich"
        4088..4216
        /rpt_family="L2"
        5232..5301
        /rpt_family="MER102"
        complement(5450..5749)
        /rpt_family="AluY"
        6571..6616
        /rpt_family="(TG)n"
        complement(6678..7320)
        /rpt_family="LIMES3"
        7419..7454
        /rpt_family="AT-rich"
        complement(7741..7807)

  misc_feature
    Query Match 50.0%; Score 689; DB 9; Length 154090;
    Best Local Similarity 80.1%; Pred.No.4e-193;
    Matches 850; Conservative 0; Mismatches 175; Indels 36; Gaps 2;

  repeat_region
    Y 351 GAATATTGGCTTAATGAGGACCAATTTTGGCTTAATAAGGAAGAAATTTCTAAGCAGCA 410
    Db 101594 GAACCTTGAACCTTGAGACACATGATTTAGGGTATCTGGGGGAGAAATTTCTAAGTAGCA 101653

  repeat_region
    Y 411 AAGCATTCAGAGGTGACTTGGGTGCTGTAAAGGCATTCAGTTTCATAAGGGAGGAGCAGA 470
    Db 101654 AAGCATTAAGAGGTGACTTAGATGGTTAAAGGTATTTCAGTTTATAAAGGGAGGAGCAGA 101713

  repeat_region
    Y 471 GCATAGAGTTTCAGAAATTTCCACCCCTGACAAATGTGATAAAGAAAGAAACCCATTTC 530
    Db 101714 GCATAACGTTTGGAAATTTTCAGCCTGACAAATGTGATAAAGAAAGAAACCCATTTC 101773

  repeat_region
    Y 531 TCAGGGGAATTCAGGTGGCTGCGAAGATTTGACATATGCTAATGAGGAGCTCAATGTTAA 590
    Db 101774 TCAGGGGAATTTGAAGCTGGCTGCGAAGATTTGACATATGCTAATGAGGAGCTCAATGTTAA 101833

  repeat_region
    Y 591 TCCTCAAGACAAATGGGAAATATATCTCTGGACATGTCAGAGGTCTTCACAGCAGTCCAT 650
```

```
Db 101834 TCCCCAAGACAAATGGGAAATGTCTCCAAGGCATATACAGAGTCTTTCATGCGAGCCCAT 101893
Qy 651 CAATACACTGGCTGGAGGCTAGGAGAAATGGTTTCTGGGACACAGGCCAGGCTTCCCT 710
Db 101894 CCATACAGGCGCGGAGGCTAGGAGAAATGGTTTCTAGGCCAGGCCAGGAGCCCC 101953
Qy 711 GTGCTGTGTGCAGCCTAGAGACTTGGTGCCCTGTGT- 746
Db 101954 ATGCTGTGTGCAGCCTAGGAGCTTGGTGCCCTGTGTATACAGCTGCTGCAGTGGTGGCT 102013
Qy 747 -----CCAGTTAACTTACGCTGTGGCTTCAGAGGGTGCAAGGCCCAAGCCCTTGG 795
Db 102014 GAAAGGGGCCAGCATAGAGCTTGGGCGGTGCTTCGAGGGTGCAAGGCCCAAGCCCTTGG 102073
Qy 796 CAGCTTCCAAAGTGTGTTGAGCCTGTGGGTGCAAGAAGCTCAAGAAATTCAGGTTTGGGAA 855
Db 102074 CAATTTCCATGTGTTGAGCCTGAGGTGCACAGAAGTCAAGAAATTCAGGCTTGGGAG 102133
Qy 856 CCTCAATCAGATTTCCAGAAGATATATGAAACCCCTGGATGCCAGGCCAGAGTTTGTCT 915
Db 102134 CCTCCACCTTAGATTTTCAGAGGATGTATGAAATGCTGGATGTCCAGGAGGAGTTTGTCT 102193
Qy 916 GTAGGGTGGGTCTCTCATGGAGAACCTCTGCAAGGTTAGTACAAAGGAAATGTTGGG 975
Db 102194 TCAGGGTGGGCGCCTCATGGAGAACCTCTGCTAGGGCAGTGCAGAAAGGAAATGTTGGG 102253
Qy 976 TGGAGGCCCCACACAGAGTCCCGAGTGGGCTCCATCTAGTAGAGCTGTGAGAGCAAGCT 1035
Db 102254 TTGGAGGCCCATACACAGAGTCCCTACTGGGGCACCCTCTAGGGGAGCTGTGAGAGAGGG 102313
Qy 1036 CCACCATCTCCAGACTCCAGAAGGTAGATCCACTGACAGCTTGCAGCATGTGCGCTGAA 1095
Db 102314 CCACCATCTCCAGACCCCGAGATGATGAAATGCTGGATGTCCAGGAGGAGTTTGTCTGCA 102373
Qy 1096 AATCCACAGACACTCAGTGCAGGCTGTGAAGAGCAGGAGTGGAGTCTGTACCCCTAC 1155
Db 102374 AAAGCCACAGAGACTCAACACCGCCCATGAAAGCAGCCCAAGAGGAGGAGCTGTACCCCTGC 102433
Qy 1156 AAACCGTAGTGCAGAGCTGACCAAGACCGTGGGAATCTACCTCTTGCATTTCTCATGAC 1215
Db 102434 AAAGCCACAGAGCAGAGCTTCCCAAGACCCATGGAACCCACTCTTGCATGAGTGTGAC 102493
Qy 1216 CTGGAGCTGAGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTTGACTGCCCC 1275
Db 102494 CTAGATGTGAACATGCACTCAAGAGATCATATTAGAATTTGAGATTTGACTACCT 102553
Qy 1276 ACTGATTTGGAGCTTATATGGGCGCGTA-CCCTTTTGTGTTGGCAATTTTTCCTCAT 1334
Db 102554 TCTGGGTTTCAGACTTGCATGGGCGCATAGCCCTTTGTTTGGCAATTTTCTCCCAT 102613
Qy 1335 TGAAGCTGCCGTATTTACCCAAATGCTGTACCTCCATTCGA 1375
Db 102614 TAGAACAGCTGTATTTACTCAATGCTGTATCGCCATTCGA 102654

RESULT 8
AC123595
LOCUS AC123595 174874 bp DNA linear PRI 01-JUN-2002
DEFINITION Homo sapiens chromosome 5 clone RP11-213H15, complete sequence.
ACCESSION AC123595
VERSION AC123595.1 GI:21306655
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 174874)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 174874)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
```





## AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castile,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Menga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE  
JOURNAL

Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 189768)

REFERENCE  
AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castile,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Menga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE  
JOURNAL

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 22, 2000 this sequence version replaced gi:7543856.

## COMMENT

All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L9594

Center clone name: 792\_C\_1

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 168749 bases at least Q40

Consensus quality: 180047 bases at least Q30

Consensus quality: 184014 bases at least Q20

Insert size: 190000; agarose-fp

Insert size: 186268; sum-of-contigs

Quality coverage: 3.8 in Q20 bases; agarose-fp

Quality coverage: 3.9 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 36 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1015: contig of 1015 bp in length  
\* 1016 1115: gap of 100 bp  
\* 1116 2284: contig of 1169 bp in length  
\* 2285 2384: gap of 100 bp  
\* 2385 3898: contig of 1514 bp in length  
\* 3899 3998: gap of 100 bp  
\* 3999 5259: contig of 1261 bp in length  
\* 5260 5359: gap of 100 bp  
\* 5360 6912: contig of 1553 bp in length  
\* 6913 7012: gap of 100 bp  
\* 7013 8145: contig of 1133 bp in length  
\* 8146 8245: gap of 100 bp  
\* 8246 9694: contig of 1449 bp in length  
\* 9695 9794: gap of 100 bp  
\* 9795 11004: contig of 1210 bp in length  
\* 11005 11104: gap of 100 bp  
\* 11105 12907: contig of 1803 bp in length  
\* 12908 13007: gap of 100 bp  
\* 13008 14916: contig of 1909 bp in length  
\* 14917 15016: gap of 100 bp  
\* 15017 17114: contig of 2098 bp in length  
\* 17115 17214: gap of 100 bp  
\* 17215 18527: contig of 1313 bp in length  
\* 18528 18627: gap of 100 bp  
\* 18628 20561: contig of 1934 bp in length  
\* 20562 20661: gap of 100 bp  
\* 20662 22484: contig of 1823 bp in length  
\* 22485 22584: gap of 100 bp  
\* 22585 25287: contig of 2703 bp in length  
\* 25288 25387: gap of 100 bp  
\* 25388 27603: contig of 2216 bp in length  
\* 27604 27703: gap of 100 bp  
\* 27704 30020: contig of 2317 bp in length  
\* 30021 30120: gap of 100 bp  
\* 30121 33472: contig of 3352 bp in length  
\* 33473 33572: gap of 100 bp  
\* 33573 38084: contig of 4512 bp in length  
\* 38085 38184: gap of 100 bp  
\* 38185 41069: contig of 2885 bp in length  
\* 41070 41169: gap of 100 bp  
\* 41170 45266: contig of 4097 bp in length  
\* 45267 45366: gap of 100 bp  
\* 45367 51092: contig of 5726 bp in length  
\* 51093 51192: gap of 100 bp  
\* 51193 55262: contig of 4070 bp in length  
\* 55263 55362: gap of 100 bp  
\* 55363 60982: contig of 5620 bp in length  
\* 60983 61082: gap of 100 bp  
\* 61083 67628: contig of 6546 bp in length  
\* 67629 67728: gap of 100 bp  
\* 67729 72690: contig of 4962 bp in length  
\* 72691 72790: gap of 100 bp  
\* 72791 79286: contig of 6496 bp in length  
\* 79287 79386: gap of 100 bp  
\* 79387 83451: contig of 4065 bp in length  
\* 83452 83551: gap of 100 bp  
\* 83552 92170: contig of 8619 bp in length  
\* 92171 92270: gap of 100 bp  
\* 92271 96333: contig of 4063 bp in length  
\* 96334 96433: gap of 100 bp  
\* 96434 103218: contig of 6785 bp in length  
\* 103219 103318: gap of 100 bp  
\* 103319 112553: contig of 9235 bp in length  
\* 112554 112653: gap of 100 bp  
\* 112654 123239: contig of 10586 bp in length



www-shgc.stanford.edu  
Quality: Phrap Quality >=40 99.8% of Sequence;  
Estimated Total Number of Errors is 0.2.  
STS Content:  
SHGC-10855 G13672  
SHGC-53350 G36743  
WI-14854 G22310  
SHGC-64132 G38710.

FEATURES  
source  
Location/Qualifiers  
1. .170368  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-54C4"  
BASE COUNT 50024 a 37145 c 34373 g 48826 t  
ORIGIN

Query Match 49.7%; Score 684.2; DB 9; Length 170368;  
Best Local Similarity 82.0%; Pred. No. 1.1e-191;  
Matches 827; Conservative 0; Mismatches 148; Indels 34; Gaps 2;  
  
QY 402 TAAGCAGCAAGCATTCAAGAGGTGACTTGGTGTCTGTAAAGGCATTCAGTTTCATAAG 461  
DB 25774 TAAGCAGCAAGCATTCAAGAGTTATGTGTTGCTGTAAAGGCACCTCGTTTATAG 25715  
  
QY 462 GGAGCAGAGCATTAAGAGTTTCAGAAATTTGCACCCTGCACATGTGATATAAAGAAAA 521  
DB 25714 GGAAGCAGAGCATATAAATTTGGAAAAATTTGCAGCCTGCACATGTGCTAGAAAGAAAA 25655  
  
QY 522 CCCATTTCGTGGGGAATTCAGCTGGCTGCAGAAATTTGCATATGTAATGAGAGCT 581  
DB 25654 CCCATTTCGTGGGGAATTTGAAGCCAGCTGAAGAAATTTTCATAAGTAATGAGAGCC 25595  
  
QY 582 GAATGTTAATCTCAAGACAATGGGAAAAATATCTCCCTGGACATGTCAGAGTCTTCACA 641  
DB 25594 AAATGTTAATCCCAAGACAATGGGAAAAATGTCTCCGAGGATGTGACAGGTCTTCACA 25535  
  
QY 642 CGAGTCCATCAATCACTGGCCTGGAGCCTAGGAGAAATGTTTGTGGACAGGCC 701  
DB 25534 CGAGCCTACTCCATCACAGGCCCAAGGCTTTGGAGAAATCGTTTCGTGGCCAGGCC 25475  
  
QY 702 AGGTCCTGTGCTGTGCACGCTAGAGACTTGGTGCCTGTGCCAGTTAATTCAGC 761  
DB 25474 AGGTCCTGTGCTGTGCTGTGCTGTGAGACTTGGTGCCTGTGCCAGCTCCAGC 25415  
  
QY 762 -----TGTGGCTTCAGAGGTCAGGCCCAA 788  
DB 25414 CATGACTAAAGGTGCCAAGTACAAATTCAGGTGTTGTTCTTCAGAGGTGGAAGCCCAA 25355  
  
QY 789 GCCTTGGCAGCTTCCAAGTGGTGTGAGCCTGTGGTGCACAAAGTCAAGAAATTCAGGT 848  
DB 25354 GCCTTGGCAGCTTCCACGTGGCATTGAGCCTGCGGTGCACAGAGTCAAGAAATTCAGT 25295  
  
QY 849 TTGGGAACCTCCANACAGATTTTCAGAGATATATGAAACCCCTGGATGCCAGGCAGAA 908  
DB 25294 TTGGTAACCTCTGCCTAGATTTTCAGAGGTGTATGAAACCCCTGGATGCCCAAGCAAA 25235  
  
QY 909 GTTTCGTGTAGGGTGGGTCTCTATGGAGAACCTCTGCAAGGCTAGTACAAAGGAAA 968  
DB 25234 GTCTGCTGAGGGGTGGGGCCCTCATGGAGAACCTCTGCTGAGGAGTGCAGAGGGAAA 25175  
  
QY 969 TGTGGTGGGAGCCCCACACAGAGTCCCCAGTGGGGCTCCATCTAGTAGAGCTGTGAG 1028  
DB 25174 TGTGGGTTTGAAGCCCCACACAGAGTCCCTACTTGGGCACCTCTAGTGGAGCTGTGAG 25115  
  
QY 1029 AGAAGTCCACCATCTCTCAGACTCCAGAGGTTAGATCCATGACAGCTTCACAGATGT 1088  
DB 25114 AAGAGCGCCCAACACACTCCAGACCCAGAAATGTTAGATCCATGACAACTTCGACTGTGA 25055  
  
QY 1089 GCTTGAATAATCCACAGACACTCAGTGCAGGCTGTGAAAGCAGCAGGAGTGGAGTCTGT 1148  
DB 25054 GCTTGAATAATCCACAGACACTCAGGCTAGCTATGAAAGCAGCAGGAGGAGGCTCT 24995

QY 1149 ACCCTACAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACTCTTGCATTG 1208  
DB 24994 ACCCTGCAAAAGCCACAGGGGCATAGCTGCCCAAGACCATGGGAACCCCTTGCATCA 24935  
  
QY 1209 TCATGACCTGGACGTGAGACATGGAGATCAAAAGAGATCATTTGGAGCTTTAAGATTGA 1268  
DB 24934 GTGTGACCTGCATATGAGACGTGGAGTCAAAAGAGATCATTTGGATCTTTAAGACTTGA 24875  
  
QY 1269 CTGCCCCACTGGATTTCGGACTTATATGGGGCCGTA-CCCCTTTTGTGGCCAATTTT 1327  
DB 24874 CTGTCTACTGGATTTCACACTCGCATGGGCTGTAGCCCTTGTGTGGCCATGTC 24815  
  
QY 1328 TTCCATTGGAACTGCCGTATTATTACCAATGCTGTACCTCCATGTAT 1376  
DB 24814 TCCCAATTAGATGGCTGTATTATTATCCAATACCTGTACCCCATGTAT 24766

RESULT 12  
AC010312  
LOCUS Homo sapiens chromosome 5 clone CTB-62P13, complete sequence.  
DEFINITION AC010312  
ACCESSION AC010312  
VERSION AC010312.5 GI:15281196  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 67984)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 67984)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE 3 (bases 1 to 67984)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (03-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE 4 (bases 1 to 67984)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (26-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE 5 (bases 1 to 67984)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (23-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT On Aug 23, 2001 this sequence version replaced gi:11079410.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www-shgc.stanford.edu  
Quality: Phrap Quality >=40 99.7% of Sequence;  
Estimated Total Number of Errors is 0.1.  
STS Content:  
SHGC-10855 G13672.

FEATURES  
source  
Location/Qualifiers  
1. .67984  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTB-62P13"  
BASE COUNT 19168 a 13195 c 14662 g 20959 t  
ORIGIN  
  
Query Match 49.6%; Score 683.6; DB 9; Length 67984;  
Best Local Similarity 81.9%; Pred. No. 1.5e-191;  
Matches 827; Conservative 0; Mismatches 149; Indels 34; Gaps 2;

Qy 401 CTAAGCAGCAAGCATTCAGAGAGTGAAGTGGTCTCTTTAAAGCATTCAGTTTCATAA 460  
Dy 7119 CTAAGCAGCAAGCATTCAGAGAGTGAAGTGGTCTCTTTAAAGCATTCAGTTTCATAA 7178  
Qy 461 GGGAGGCAGACATTAAGAGTTCAGAAAATTTGCCACCTTCGACATGTGATAAAAAAGAAA 520  
Dy 7179 GGGAGGCAGACATTAAGAGTTCAGAAAATTTGCCACCTTCGACATGTGATAAAAAAGAAA 7238  
Qy 521 ACCATTTTCTGAGGGGAAATTCAGCTGGCTGCGCAGAAAATTTGCAATGTGATGAGGAGC 580  
Dy 7239 ACCATTTTCTGAGGGGAAATTTGAGCCAGCTGAGAAAATTTGCAATGTGATGAGGAGC 7298  
Qy 581 TGAATGTTAACTCAAGACAAATGGGAAATATCTCTGACATGTGACAGAGTCTTCAC 640  
Dy 7299 CAAATGTTAACTCAAGACAAATGGGAAATATCTCTGACAGTGTGACAGAGTCTTCAC 7358  
Qy 641 AGCAGTCCATCAATCACTGCGCTGGAGCCCTAGGAGAAAATTTGTTGTGGGACAGGCC 700  
Dy 7359 AGCAGTCCATCAATCACTGCGCTGGAGCCCTAGGAGAAAATTTGTTGTGGGACAGGCC 7418  
Qy 701 CAGGTCCTCTGTCTGTGCTGAGCCTAGAGACTTGGTGCCTGTGCTCCAGTTAATTCAG 760  
Dy 7419 CAGGTCCTCTGTCTGTGCTGAGCCTAGAGACTTGGTGCCTGTGCTCCAGTTCAG 7478  
Qy 761 C-----TGTTGGCTTCAGAGGGTGCAGGCCCA 787  
Dy 7479 CCATGACTAAAAGGTGCCAAAGTACAATTCAGTTGTTCTTCAGAGGTGGAAGGCCCA 7538  
Qy 788 AGCCTTGGCAGCTTCCAAGTGGTGTGAGCCTGTGGGTGCGAAAGTCAAGATTGAGG 847  
Dy 7539 AGCCTTGGCAGCTTCCAAGTGGTGTGAGCCTGTGGGTGCGAAAGTCAAGATTGAGG 7598  
Qy 848 TTTGGAACTCCCAATCAGATTTTCAGAGATATATGGAACCCCTGGATGCCAGGCAGA 907  
Dy 7599 TTTGGTAACCTTCGCTAGATTTTCAGAGAGTGTATGGAACCCCTGGATGCCAGGCAGA 7658  
Qy 908 AGTTTGTGTAGGGTGGGTGCTCATGAGAACTCTGCAAGGCTAGTACAAAAGGAA 967  
Dy 7659 AGTCTGCTGCAAGGGTGGGCGCTCATGAGAACTCTGCTAGGCGAGTGCAGAAAGGAA 7718  
Qy 968 ATGTTGGTGGAGGCCCCACACAGAGTCCCAGTGGGGCTCCATCTAGTAGAGCTGTGA 1027  
Dy 7719 ATGTTGGGTTGAAGGCCCCACACAGAGTCCCCTACTTGGGCACTCTAGTAGAGCTGTGA 7778  
Qy 1028 GAAGAAGTCCACCTCCAGACTCCAGAGGTTAGATCCACTGACAGCTTGCAGCATG 1087  
Dy 7779 GAAGAGGCCACCACTCCAGACCCCAAGATGTTAGATCCACTGACAGCTTGCAGCATG 7838  
Qy 1088 TGCCTGAAAAATCCACAGACTCAGTCCAGCCTGTGAAAGCAGCAGGAGTGGAGTGTG 1147  
Dy 7839 AGCCTGGAAGAACCCACAGACTCAACGCTAGCCTATGAAAGCAGCAGGAGGAGGCTC 7898  
Qy 1148 TACCTACAAAACCGTAGTGGCAGCTGACCAAGACCGTGGGATCTACCTCTTGCATT 1207  
Dy 7899 TACCTGCAAGGCCACAGGGGCTAGCTGCCCCAAGACCATGGAACCCCTTGCATC 7958  
Qy 1208 GTCATGACCTGACGTGAGCATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTG 1267  
Dy 7959 AGTGGACCTGCATATGAGACGTGGAGTCAAAAGAGATCATTTTGGATCTTTAGACTTG 8018  
Qy 1268 ACTGCCCCACTGGATTTTCGGACTATATATGGGGCCCGTA-CCCCCTTTGTTTGGCCAAATTT 1326  
Dy 8019 ACTGCTCTACTGGATTTTCAGACTGCGATGGGCGCTGTAGCCCCCTTTGTTTGGCCAAATGT 8078  
Qy 1327 TTTCCATTTGGAATGCGCTATTTTACCAATGCGCTGTACCTCCATGTAT 1376  
Dy 8079 CTCCCATTTAGAAATGGCTGTATTTTATCAATACCTGTACCCGCCCATTTGTAT 8128

RESULT 13  
AC091005/c  
LOCUS

DEFINITION  
AC091005  
VERSION  
AC091005.9 GI:17488655  
KEYWORDS  
HTG.  
SOURCE  
Homo sapiens.  
ORGANISM  
Homo sapiens

REFERENCE  
1 (bases 1 to 168502)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 15, clone RP11-1008C21  
Unpublished

TITLE  
JOURNAL  
AUTHORS  
REFERENCE

TITLE  
JOURNAL  
AUTHORS  
REFERENCE

TITLE  
JOURNAL  
AUTHORS  
REFERENCE

Homo sapiens chromosome 15, clone RP11-1008C21, complete sequence.  
AC091005  
AC091005.9 GI:17488655  
HTG.  
Homo sapiens.  
Homo sapiens  
1 (bases 1 to 168502)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 15, clone RP11-1008C21  
Unpublished

2 (bases 1 to 168502)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,  
Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,  
Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K.,  
Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,  
Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,  
McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,  
Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,  
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,  
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,  
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,  
Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,  
Theodore,J., Travers,M., Travis,N., Triggilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (24-MAR-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 168502)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,  
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,A.,  
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,  
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Triggilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (01-SEP-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 168502)  
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,  
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,A.,  
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,

Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacClean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Plerre, N., Pollard, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (11-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Dec 11, 2001 this sequence version replaced gi:15412474.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L10835  
Center clone name: 1008\_C\_21  
-----

FEATURES  
source Location/Qualifiers  
1. .168502  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/map="15"  
/clone="RP11-1008C21"  
/clone\_lib="RPC1-11 Human Male BAC"  
140. .191  
/rpt\_family="L2"  
194. .218  
/rpt\_family="(CAA)n"  
complement(222. .381)  
/rpt\_family="FRAM"  
384. .464  
/rpt\_family="L2"  
412. .417  
/note="<30 qual SNGL region"  
667. .672  
/note="<30 qual SNGL region"  
799. .803  
/note="<30 qual SNGL region"  
1130. .1158  
/rpt\_family="(CA)n"  
complement(1886. .1985)  
/rpt\_family="MIR"  
complement(2627. .2637)  
/note="<30 qual SNGL region"  
3489. .3530  
/rpt\_family="(A)n"  
4319. .4348  
/rpt\_family="AT\_rich"  
4775. .4903  
/rpt\_family="MER85"  
5137. .5175  
/rpt\_family="AT\_rich"  
5198. .5557  
/rpt\_family="MLT2F"  
5558. .5595  
/rpt\_family="(CA)n"  
5596. .5624  
/rpt\_family="MLT2F"  
complement(6251. .6387)  
/rpt\_family="MIR"

Qy	591	TCCTCAAGCAATGGGAAAAATATCTCTCTGGACATGTCAGAGGTCTTCACAGCATGTCAT	650
Db	45484	TCCCCAAGACAATGGGAAAAATGTCTCCAGGCATGTTCAGAGGTCTTCATAACAGCCCT	45425
Qy	651	CAAAATCACTGGCGCTGAGGCGCTAGAGAAATGGTTTGTGGGACAGGCCAGGTCCT	710
Db	45424	GCCATCAAGGCGCTGGAGGCGCTAGAGAAATGGTTTGTGGGACAGGCCAGGTCCT	45365
Qy	711	GTGCTGTGTGACGCTAGAGACTTGGTCCCTGTGTCCAGTTAATTCAG-----	760
Db	45364	ATGCTGTGTGACGCTAGGACTTGTTCCTCTGCATCCAGCTGCTCAGCCATGGCTGA	45305
Qy	761	-----CTGTGGCTTCAGAGGGTGCAGCCCAAGCCCAAGCCCTTGGCA	797
Db	45304	AAGGGCCCAACATAGAGCTCGGACTGTGGCTTCAGAGGTGGAAGCCCAAGCCCTTGA	45245
Qy	798	GCTTCCAAGTGGTGTGAGCCTGTGGGTGCAAGAGTCAAGAATTGAGGTTTGGGAACC	857
Db	45244	GCTTCCAGCTGGTGTGAGCCTTACAGGTGCACAGAAGTCAAGAATGGGTTTGGGAACC	45185
Qy	858	TCCAAATCAGATTTCAGAGATATATGAAACCCCTGGATGCCAGGCAGAAATTTGCTGT	917
Db	45184	TCCGCTAGATTTTAGAAGATGTATAGAAACACCTGGATGCCAGGCAGAAATTTGCTGT	45125
Qy	918	AGGGTGGGCTCCTCATCGAGAACCTCTGCAAGGCTAGTACAAAGGGAATTTGGGTG	977
Db	45124	AGGGTGGGAGTTCATGGAGAACCTCTACTAGGCGACATGGGAAGGGAATTTGGGTG	45065
Qy	978	GGAGCCCCACACAGAGTCCCCAGTGGGCTCCATCTAGTAGAGCTGTGAGAAGAGTCC	1037
Db	45064	AGATCCCCCACACAGAGTCCCTCTCTGGGGCACGTGCTAGTGGAGCTGTGAGAAGAGGCC	45005
Qy	1038	ACCATCTCCAGACTCCAGAGGCTAGATCCACTGACAGCTTGACAGCATGTGCCTGAAAA	1097
Db	45004	ACTGCTCTCCAGACCCAGATGGTAGATCCACGACAGCTTGACCATTTGCCTGGAAA	44945
Qy	1098	ATCCACAGACACTAGTCCGACCTGTGAAAGCAGCAGGAGTGGAGTCTGTACCCCTACAA	1157
Db	44944	AACGGCAGACACTCAATTCAGCCCGCAGGAAAGCAGCCAGGAGGAGGTATACCCCTGCAA	44885
Qy	1158	AACGCTAGTCAGAGCTGACCAAGACCGTGGATCTACCTCTGCAATTTGTCATGACCT	1217
Db	44884	AACCAAGGGTGGAGTGGCCCAAGACCATCTTTGGACCCACCTCTTGCATCAGCATGACCT	44825
Qy	1218	GGAGCTGAGACATGAGTCAAAAGAGATCATTTTGGAGCTTTTAAGATTTGACTGCCCCAC	1277
Db	44824	GGATGTGAGACCTGGAGTCAAGAGAGATCATTTGGAGCTTTTAAATTTGACTGCCCCAC	44765
Qy	1278	TGGATTTCCGACTTATATATGGGCGCGT-ACCCCTTTGTTTGGCCAAATTTTTCATTTG	1336
Db	44764	TGGATTTGGGACTTGCATGGGCGCGTGAACCCCTTTGTTTAGGCCAAATTTCTCCCATTTG	44705
Qy	1337	GAAGTGGCGTATTTACCAATGCTGTACCTCCATTTGAT	1376
Db	44704	GAATGGCTGATTTTACCAATACCTGTACTGCTCATTTGAT	44665
RESULT 14			
AC079614			
LOCUS	AC079614	131215 bp	DNA linear PRI 09-JAN-2002
DEFINITION	Homo sapiens BAC clone RP11-42414 from 2, complete sequence.		
ACCESSION	AC079614		
VERSION	AC079614.7 GI:15431230		
KEYWORDS	HTG.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 131215)		
AUTHORS	Sulston, J.E. and Waterston, R.		
TITLE	Toward a complete human genome sequence		
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)		

99063792	MEDLINE
9847074	PUBMED
2 (bases 1 to 131215)	REFERENCE
Nguyen, C., Abbott, A. and Elliott, G.	AUTHORS
The sequence of Homo sapiens BAC clone RP11-42414	TITLE
Unpublished (2001)	JOURNAL
3 (bases 1 to 131215)	REFERENCE
Waterston, R.H.	AUTHORS
Direct Submission	TITLE
Submitted (04-SEP-2000) Genome Sequencing Center, Washington	JOURNAL
University School of Medicine, 4444 Forest Park Parkway, St. Louis,	
MO 63108, USA	
4 (bases 1 to 131215)	REFERENCE
Waterston, R.H.	AUTHORS
Direct Submission	TITLE
Submitted (05-SEP-2001) Genome Sequencing Center, Washington	JOURNAL
University School of Medicine, 4444 Forest Park Parkway, St. Louis,	
MO 63108, USA	
5 (bases 1 to 131215)	REFERENCE
Waterston, R.	AUTHORS
Direct Submission	TITLE
Submitted (09-JAN-2002) Department of Genetics, Washington	JOURNAL
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA	
On Sep 5, 2001 this sequence version replaced gi:15209227.	COMMENT
-----	
Genome Center	
Center: Washington University Genome Sequencing Center	
Center code: WUGSC	
Web site: http://genome.wustl.edu/gsc	
Contact: sapiens@watson.wustl.edu	
-----	
Summary Statistics	
-----	
Center project name: H_NH0424104	
-----	

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:  
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Taten, M., Catanese, J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RP11-22L19, 2000 bp overlap; the clone sequenced to the right is RP11-364I15. Actual start of this clone is at base position 128235 of RP11-22L19; actual end is at base position 131215 of RP11-42414.

The sequence between 58007 and 58074 is covered only by a PCR product of clone DNA; the sequence between 58180 and 58221 is covered only by a PCR product of clone DNA. Data from AC093650 was used to finish this clone, AC079614. Polymorphisms exist between RP11-42414 and RP11-364I15.



```

QY 785 CCAAGCCTTGGCAGCTTCCAAAGTGGTGTGAGCCCTGGTGGTGAAGAAAGTCAAGATTG 844
|||||
Db 128705 TCAAGCCTTGGCAGCTTCCATGTGTGTGAGCCCTGCAAGTGCACAGAACTCAAGATTG 128764
QY 845 AGGTTTGGGAACCTTCCAATCAGATTTTCAGAGATATATGGAACCCCTGGATGCCCGCC 904
|||||
Db 128765 AGGTTTGGGAACCTTCCCTAGATTTTCAGAGATATATGGAACCCCTGGATGCCCGCC 128824
QY 905 AGAAGTTTGGTGTAGGGTGGGCTCTCATGAGAGACTCTGCAAGGGTGTAGTACAAAAG 964
|||||
Db 128825 AGAAGTTTGGTGTAGGGTGGGCTCTCATGAGAGACTCTGCTAGGCAAGTGTGAAG 128884
QY 965 GAAATGTTGGTGGAGAGCCGCCACACAGAGTCCCAAGTGGGGTCCATCTAGTAGAGCTG 1024
|||||
Db 128885 GAAATGTTGGTGGAGAGCTCCACACAGAGTCCCTACTGGGGCCACCACCTAATGGAGCTG 128944
QY 1025 TGACAAGAAGTCCACCATCTCCAGACTCCAGAGAGGTAGATCCACTGCACAGCTTGCAGC 1084
|||||
Db 128945 TGACAAGAAGTCCCAAGTCCAGAGAGGTAGATCCACTGCACAGCTTGCAGC 129004
QY 1085 ATGTGCTCTGAAAATCCAC--AGACACTCAGTGCAGCCTGTGAAAGCAGCAGCGATGGA 1142
|||||
Db 129005 ATGTGCTCAGGAAGACCCACAGAGAGTCAATGCCAGCCCATGAAGCAGCGATGGA 129064
QY 1143 GTCTGTACCTTCAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTCTT 1202
|||||
Db 129065 GGCTGTGACATTTCAAAGCCACAGCGGAGAGTGTGCCAAGACCGCGGACCCACCTCTT 129124
QY 1203 GCATTGTGATGACCTGACGTGACACATGGATGGAATCAAAAGAGATCATTTTGGAGCTTAA 1262
|||||
Db 129125 GCATTGTGACCTGACGTGACACATGGATGGAATCAAAAGAGATCATTTTGGAGCTTAA 129184
QY 1263 ATTTGACTGCCCTTGGATTTCCGACTTATATGGGCGCCGT-ACCCCTTTGTTTGGCC 1321
|||||
Db 129185 ATTTGACTGCCCTTGGATTTCCGACTTATATGGGCGCCGT-ACCCCTTTGTTTGGCC 129244
QY 1322 AATTTTTCATTTGGAACGTCGCTATTTACCCCAATGCCT 1361
|||||
Db 129245 AACATCTCCATTTGGAATGGCTGTATTTACCCCAATGCCT 129284

```

## RESULT 15

AC012038/c

LOCUS

DEFINITION Homo sapiens 12 BAC RP11-547C5 (Roswell Park Cancer Institute Human BAC Library) complete sequence.

AC012038

AC012038 19 GI:12000459

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hughes,M., Holloway,C., Hollins,B., Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,

Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvan,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseg,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S., Metzger,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwunonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.D., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shlim,C., Shooshitari,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,J.R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 73390)

Worley,K.C.

Direct Submission

Submitted (19-OCT-1999)

Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 73390)

Worley,K.C.

Direct Submission

Submitted (01-JAN-2001)

Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Jan 1, 2001 this sequence version replaced gi:11138159.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email

gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the

entire insert of this clone. Overlapping regions of clones are only

sequenced and submitted once, so the sequence for the remainder of

the insert may be found in the record for the adjacent clones.

Overlapping clones are noted at the beginning and end of the

Features listing.

FEATURES

STSS are identified using ePCR (Genome Res. 7:541-550) searches

of a local database that includes entries from dbSTS, GDB, and

local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,

unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST

(Nuc. Acids Res. 25:3389-3402) similarity (expect &lt; 1e-34) to the

EST and cDNA sequences. Genes demonstrate at least two exons

flanked by consensus splice sites that maintained sequence

continuity across the splice junctions. Sequences that are not

identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum

standard of double strand coverage with a minimum of 2 clones and 2

reads with no ambiguities or 2 chemistries with a minimum of 2

clones and 3 reads with no ambiguities. If the sequence quality for

a region does not meet this standard, it will be indicated in the

annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality

standards - estimated error rate less than 1 per 10,000 bases.

Reports of lowest quality individual bases and measures of base

quality are listed below. Description of the metrics can be found

at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation>.





THIS PAGE BLANK (USPTO)